MORALITY, DIGNITY
AND PRAGMATISM

AN ESSAY ON THE FUTURE OF MORALITY

JAMES GEORGE SCOTT WILSON

University College London

Submitted for the examination of PhD, January 2002

Abstract of *Morality, Dignity and Pragmatism*

This thesis is an examination and reconstruction of morality. It divides into three parts.

Part one argues that morality is best considered as the tradition of ethical thinking that begins with the Stoics, develops in Christian thought and reaches its apotheosis in Kant. This tradition structures ethical thinking around three basic concepts: *cosmopolitanism,* or universal applicability to human beings as such, the *dignity of human beings* and *reciprocity.* It is this tradition of morality that Nietzsche sets out to destroy.

Part one criticises pre-Nietzschean theories of morality, such as Kant's, that take universal and exceptionless rules to form the core of morality. It critiques both the possibility of putting forward an adequate set of such rules and the proposed relationship between morality and human life that is implicit in these theories.

Part two begins with Nietzsche's challenge: that morality is a system of values rooted in nihilistic resentment at the vitality of other, stronger modes of living. It argues that this challenge must be taken seriously, and that the best way to do this is to make it clear that morality has as its fundamental basis a responsiveness to the value of human life; hence it is Nietzsche's ethics that should be called nihilistic.

The rest of part two examines the possibility of answering Nietzsche's challenge by demonstrating a necessary connection between human selfhood and the acknowledgement of the dignity of human beings. Here I criticise Christine Korsgaard's arguments and consider Charles Taylor's more promising approach to the self.

Part three turns towards pragmatism, and in so doing gives up on the attempt to show that morality is somehow *necessary* for all human beings.

Nietzsche's challenge is answered more subtly: an empirically backed theory of human selfhood explains the point of morality in terms of our basic need for recognition.

I complete the reconstruction of morality by reinterpreting the dignity of human beings in a naturalistic way and adopting a conception of moral rules that is informed by Jürgen Habermas' discourse ethics.

For my parents

ACKNOWLEDGMENTS

CONTENTS

PREFACE

This thesis is a constructive work in the tradition of morality.

Although the argument is long and complex, the aims are simple: to explain the point and value of goodness without straying into wishful thinking or laughable metaphysics; to maintain the idea of moral duty whilst frankly admitting that the heart of morality is elsewhere; and to provide an honest document that may actually be of some use to those who read it in the conduct of their lives.

It is also in large part an essay about the meaning of Kant for moral philosophy. Although I see no good reason why this should be so, my position on Kant appears to be somewhat unusual, combining as it does a profound respect and admiration for his contribution to moral philosophy with a series of searching and rather brutal criticisms of it.

I can only say that this has always seemed to me to be the best way to do philosophy: to take the work of those whom one admires most and submit it to a thorough critique, in the hope that one can build something better on the foundations that they have laid.

NOTE ON REFERENCES AND ABBREVIATIONS

References to Kant's texts are abbreviated as follows: *G – Groundwork to the Metaphysics of Morals; C1 – Critique of Pure Reason; C2 – Critique of Practical Reason; C3 – Critique of Judgment; MM – Metaphysics of Morals; PP – Toward Perpetual Peace: A Philosophical Sketch; R– Religion within the limits of reason alone.* All page references are to the Akadamie edition, except in the case of the *Religion* and *Perpetual Peace,* where the translations used do not carry the Akadamie page numbers. Here references are to pages in the translations used.

References to Nietzsche are to section numbers rather than pages. References to Aristotle are to the standard Bekker page numbers. Other classics (e.g. Seneca) are referenced by book and chapter number.

Where an author's articles have been collected into a book form, and that book is also cited in the bibliography, page references will be to the book reprint.

PART ONE:

TRADITIONAL ACCOUNTS

OF MORALITY

MAP OF PART ONE

Chapter one, *The Very Idea of Morality,* aims to define the basic territory of morality. It begins by distinguishing the usage of 'morality' adopted in this thesis from three others that are current in philosophy and discusses briefly how the concerns manifested in these usages of morality are taken up in the rest of the thesis. Next, it fills out the usage of morality adopted, both by putting it in a historical context and by discussing its central ideas of cosmopolitanism, reciprocity and human dignity. Finally, it briefly addresses the variety of positions *within* morality, and the end state at which morality aims.

Chapter two, *Against Kant, for the sake of Morality,* is a critique of Kant's account of morality, insofar as it makes the purity of morality central. It argues that i) there is no reason to follow Kant's purifying assumptions about the possibility of morality; for morality is quite possible without these assumptions. ii) Moral obligation is in fact *impossible* if we take Kant's presuppositions about the moral law seriously; iii) Kant's account of morality gets the basic structure of human life so wrong that it renders itself inappropriate for the general regulation of human behaviour.

C h a p t e r   O n e
The Very Idea of Morality

In this thesis, I shall use 'morality' to name the *tradition* in Western ethical thinking that is inaugurated by the Stoics and centres on *universalistic interpersonal justice.* What marks this tradition out from all others is i) its account of what *constitutes* interpersonal justice, and ii) the seriousness with which it attempts to reconstruct the world as it finds it in its own image. Put otherwise, to bring it about that the world *does* operate according to moral rules.

i) While morality follows the familiar and banal underlying idea of justice ('to give to each what is due to them'), it brings three inter-related ideas into play to fill out what it takes to *constitute* giving each what is due to them. These ideas are:

Cosmopolitanism - the idea that we are all citizens of the world, and thus should always at least treat one another as fellow citizens.

Reciprocity - the idea that there should be always be a *reversibility* in action, so that before A acts, he should ask himself if he would be happy for B to act in that way to him (or, more generally, if he would be happy for everyone to act in that way).

Human dignity - the idea that human beings possess a *worth* that both sets them above things, plants and other animals, and makes them equal to one another, and that this dignity must be respected at all times.

ii) Morality believes that *it matters* that, in the world as it stands, people do not get what is due to them as human beings; and it believes – stubbornly believes – that we can make progress towards a situation in which people do, generally, get what is due to them as human beings. Living and acting morally is to take up the challenge that the world as we find it offers and

impel it in the right direction. (This is composed of two parts; first, an anger – a rage – against the world as it now is, and second, a hope – a yearning – for the world as it might be.)

This usage diverges quite considerably from others in use in the philosophical literature. The first half of this chapter explains the relationship of morality as I shall use it to three other usages of 'morality'; the second half gives a further explication of my usage of 'morality'.

There are three other usages of 'morality' that it will prove helpful to compare and contrast with my own. They are as follows:

i) 'Morality' as devoid of normative connotations: on this usage, a morality is nothing more than the mores or social customs that happen to obtain *de facto* in a given society, and there will be as many moralities as there are systems of mores. No judgment is implied about the value of any of these moralities.

ii) 'Morality' as consequentialists understand it. These accounts i) separate the right (that is, the procedure for working out what is to be done) and the good (that is, the account of what is intrinsically valuable); ii) define the right as the attempt to promote the good impartially.

iii) 'Morality' as Nietzsche uses it. On this usage, a morality is a table of values to be lived by that may be either individual or society-wide. Nietzsche argues that these tables of values – moralities – should be ranked by the quality of the life that they express and/or make possible. On Nietzsche's view, morality in my sense should first be demoted to just one morality amongst many; then, sloughed off, after it fails the test by which moralities should be measured.

1.1 *My usage of 'morality' as compared to the others I have canvassed*

i) Non-normative usage

The first (non-normative) usage of 'morality' differs in three chief ways from mine; first, it will let *any* system of mores count as a morality, whilst mine only includes those standards that aim at justice understood in terms of cosmopolitanism, reciprocity and human dignity. Second, even within the one tradition that I have picked out, for the non-normative usage there will be *many* moralities. What the non-normative account would see as many, my account sees as unified by the common search for universalistic justice. Third, the non-normative usage brackets the purported *normativity* of the standards of the 'moralities' it looks at; my account takes the purported normativity of moral standards as key to understanding them.

On my usage morality is a living *tradition* in the sense that Alasdair MacIntyre uses the term, that is, 'an historically extended, socially embodied argument, and an argument precisely in part about the goods which constitute that tradition.' (1981, p. 222) The argument, of course, concerns the meaning, defensibility and practicality of a cosmopolitan conception of justice. This tradition has a unity invisible to the non-normative conception of morality, which can only see an endless play of different figurations. Those inside the tradition see in these same figurations a *narrative,* in which themes are introduced, developed, dropped then re-introduced; they are able to see what comes later as an advance or a decline, based on their understanding of cosmopolitan justice.

ii) Consequentialism

Consequentialism has much more in common with my conception of morality: both will agree only one morality, and that it lays down standards

appropriate to human conduct as such. Both will agree that these standards take everyone into consideration in some respect.

They differ in two chief ways: my usage centres around *justice,* whilst the teleological account centres around the promotion of the good; my usage takes morality to be *internal* to the goal of living one's life as a human being, whereas consequentialism takes morality to be external. The first concerns the *content* of morality; the second its relationship to selfhood and to being a human being.

The first is of less importance than the second. For, the emphasis on interpersonal justice versus the emphasis on the pursuit of the good does not necessarily lead to a difference in what one thinks should be done in given circumstances; nor in the reasons that one gives in support of one's case. For example, a value pluralist consequentialist such as Jonathan Glover has very similar normative views (and counter-factual normative views) to the ones I would associate with morality in my sense.[1]

The important difference is the second: that of the relationship between morality and human life. This difference is best expressed in Aristotelian terminology. We can distinguish between two different types of action: those in which one acts to produce something separate from that activity (*poiesis*), as when a man makes a cabinet; and those in which one acts merely for the sake of the activity itself (*praxis*), for example laughing.[2] Where a *poiesis* aims at some further end, a *praxis* is its own end. Using this distinction, we could say that consequentialism takes morality to be a *poiesis*, with the aim of producing as much good as possible*;* morality on my understanding takes it to be a *praxis*, the activity, broadly speaking, of treating other human beings justly.

---

[1] See, for example Glover (1977) and (1999).
[2] I am here ignoring cases where laughing is not a praxis, where one laughs in order to bring about a further end (such as making someone appear funny).

Where consequentialism suggests that there are certain ends that should be pursued in action, morality (on my usage of it) takes it that there are certain ends that should be *implicit in* or *constitutive of* action. For consequentialism, the ends of morality are external, and the question is how best to promote them. For morality the end is *internal,* and the question is the shape that this end imposes on human life. At this level of abstraction, the difference amounts to this: morality on my understanding of it rules out a priori the possibility of doing what is morally prohibited, even *in order to* bring about a good result; this consideration does not apply to consequentialism as it does not have a criterion of morality *external to consequences.*

Consequentialism understands the moral point of view in *impersonal* terms: as a vantage point one adopts in order to see how the good is to be impartially promoted. Hence for the consequentialist, the moral point of view needs to be brought back into dialogue with the personal point of view, the point of view from which we live. This is a result of the fact that it considers morality to be an *external check* on the personal point of view. Morality, on my understanding of it, takes the moral point of view to be *implicit in* or *constitutive of* human decision making in general, and hence experiences no difficulty in relating the moral point of view to the personal point of view.

However, no one should take the above as an attempt at an argument against consequentialism: it is not the business of this thesis to attempt to provide any arguments against consequentialism, or indeed to consider it at all. We are simply exploring a different tradition in ethical thinking: looking at its basic structure, the limitations it has suffered from through its history, and how to reconstruct it in such a way that it becomes believable for us: I mention consequentialism merely so that we can put it to one side.

iii) Nietzsche

Nietzsche's approach to morality is much more relevant to the project of this thesis. Nietzsche is a pervasive presence throughout this work, often hovering in the background, even where he is not explicitly mentioned. I shall comment briefly on Nietzsche's relationship to the project of this thesis, and then consider his account of morality in a brief and preliminary manner.

Nietzsche's greatness and his usefulness for our project of reconstructing morality lie in his astonishing skill as an interpreter. His training as a philologist leads him to treat pretty much everything – people, ethical systems, religions, science, his own life – as texts to be interpreted and, more importantly *re-evaluated* through the act of interpretation. This applies particularly to morality, which occupies more and more of his attention as his career as a writer hastens towards its unfortunate end.

His importance to the project of this thesis is immense: I have found both the content of his interpretations of morality and the underlying approach to interpretation to be of great help, though in somewhat different ways. His interpretations of morality show more clearly than anywhere else what is wrong with the tradition of morality as it reaches its apotheosis in Kant. But the approach to interpretation is one of the major sources of inspiration for the *reconstruction* of morality that is undertaken through the course of this thesis. In a word, we use methods of interpreting borrowed from Nietzsche to *revivify* morality, reinterpreting it piece by piece until we have constructed a system of ethical thought we are happy with.[3]

---

[3]However, it is clear that Nietzsche himself would be far from happy with the way I have used motifs from his method of interpreting to render plausible a system of ethical thinking that places mutual respect and the dignity of each and every human being at the centre of ethical thinking. For Nietzsche, such an ethical system is a sign of decadence and ultimately of an unhealthy approach to life.

However, I argue in chapter six that Nietzsche is just wrong here: what leads him to this conclusion is just 'an audacious generalization on the basis of very narrow, very personal, very human, all too human facts' (1886, Preface). His conclusions have no general validity, rather they point up his own particular weaknesses and inadequacies.

Nietzsche's thought on morality can best be apprehended as working in two directions simultaneously. First, it attempts to ridicule the account morality gives of itself, by showing this account to be partial, based on inversions of cause and effect and shot through with bad metaphysics: in short, incoherent as an interpretation of the world. Second, it attempts to insinuate a new interpretation of human life, that makes morality seem coherent again, but which recontextualises morality in such a way that it loses all its attractiveness.[4]

We can see these two processes going on quite clearly in Nietzsche's account of what morality is. He begins by noting that the tradition just assumes that that there is only one morality, and that philosophy's task is to furnish its rational ground. He then proceeds to ridicule these assumptions:

> Philosophers one and all have, with a strait-laced seriousness that provokes laughter ... wanted to furnish the *rational ground* of morality – and every philosopher hitherto has believed he has furnished this rational ground; morality itself, however, was taken as "given". (1886, §186)

Against this, Nietzsche insinuates an interpretation according to which there are *many* moralities – for on this new interpretation a morality is just a table of values – and which makes the fundamental problem of moral philosophy seem quite different:

> One should, in all strictness, admit *what* will be needful here for a long time to come, *what* alone is provisionally justified here: assembly of material, conceptual comprehension and arrangement of a vast domain of delicate value-feelings and value-distinctions which live, grow, beget and perish – and perhaps attempts to display the more frequent and recurring forms of these living crystallizations – as preparation of a *typology* of morals. (1886, §186)

---

[4] In this new context, for example, signs of the triumph of morality, such as the growth of democracy and the flourishing of human rights discourse seem like signs of society's decline, rather than, as most people might assume, signs of progress.

This new interpretation invites two further questions: first, why is the tradition of morality so keen to deny that it is just one ethical system amongst others? And second, what is the right way to construct our typology of ethical systems? Nietzsche's reply to these questions will take us deep into his theory of master and slave moralities: we shall take it up in chapter three.

1.2 *An introduction to morality*

Now it is time to give a fuller exposition of morality.[5] I have already stated that morality takes itself to be about giving human beings what is due to them, and that it understands what is due to human beings in terms of *cosmopolitanism, reciprocity* and *human dignity.* I shall discuss each of these in a little more depth; then consider some of the explanations that morality puts forward of why people on occasion fail to do what (morally speaking) they ought to; then consider the variety that is possible within the tradition of morality; and finally, the end at which morality aims.

Cosmopolitanism

Cosmopolitanism proclaims that each person is a citizen of the world, and hence every person should treat everyone else as if they were fellow citizens. Historically speaking, the most important formulations of the bare idea of cosmopolitanism[6] have been in terms of *natural law* and *circles of concern.*

---

[5] Henceforth I shall use 'morality' in my own sense.

[6] We shall see below how the ideas of reciprocity and human dignity can also lead us very strongly towards cosmopolitanism.

The Stoics argue for cosmopolitanism on the basis of natural law[7] as follows:

only man of all the animals has reason;[8] reason is *divine*;[9] and reason is common to all human beings.[10] Putting all these together, we get the following:

> Since, then, there is nothing better than reason, and reason is present in both man and God, there is a primordial partnership in reason between man and God. But those who share reason also share right reason; and since that is law, we men must also be thought of as partners with the gods in law. Furthermore, those who share law share justice. Now those who share all these things must be regarded as belonging to the same state. (Cicero, *De Legibus* I.23)

Hence it follows that there must be a law applicable to all men and gods; a natural law that is cosmopolitan in form. The result will be an enlarged view on the world.[11]

Hierocles, in another Stoic argument, suggests a different approach; he suggests that the trick is to stop thinking of others as so distant from us. On his model, we should think of ourselves as surrounded by a series of

---

[7] The classic definition of natural law is Cicero's: 'True law is right reason in agreement with nature; it is of universal application, unchanging and everlasting; it summons to duty by its commands, and averts from wrong doing by its prohibitions. And it does not lay its commands or prohibitions upon good men in vain, though neither have any effect on the wicked. We cannot be freed from its obligations by senate or people, and we need not look outside ourselves for an expounder or interpreter of it. And there will not be different laws at Rome and at Athens, or different laws now and in the future, but one eternal and unchangeable law will be valid for all nations and times, and there will be one master and ruler, that is, God, over us all, for he is the author of this law, its promulgator and its enforcing judge. Whoever is disobedient is fleeing from himself and denying his human nature, and by reason of this very fact he will suffer the worst penalties, even if he escapes what is commonly considered punishment.' (Cicero, *De Re Publica*, III.23)

[8] 'Of all the types and species of living creatures he is the only one that participates in reason and reflection, whereas none of the others do'. (Cicero)

[9] Cicero asks, 'what is there, I will not say in man, but in the whole of heaven and earth, more divine than reason?' Seneca puts it thus: 'What is best in man? Reason: with this he precedes the animals and follows the gods.' (*Letters* 76.9; Long & Sedley 1987, p.395) Note how closely tied this notion of the divinity of reason is to that of the dignity of human beings.

[10] 'Reason in fact – the one thing in which we are superior to the beasts ... that certainly is common to all.' Cicero, *De Legibus* I.30

[11] In Seneca's words: 'Let us take hold of the fact that there are two communities – the one, which is truly great and truly common, embracing gods and men, in which we look neither to this corner nor to that, but measure the boundaries of our state by the sun; the other, the one to which we have been assigned by our birth.' (Seneca, *De Otio*)

concentric circles; the first encompassing our body, the second our immediate family, the third, more distant relatives, the fourth our fellow tribesmen, then our fellow countrymen, and the next, the whole human race. On his view, what is required is that we keep drawing people towards the centre; both by transferring individual people from outer circles into more inward ones, and by moving the circles *themselves* towards us.[12] The appeal of this model is that it is a gradualist one: it can understand cosmopolitanism as part of a process of trying to bring other people closer to one.[13]

Reciprocity

The key idea of reciprocity is *reversibility* in action; that it is due to others to treat them as you would agree to were the tables turned. The simplest and most influential formulation of the central idea of reciprocity is the golden rule. There are two canonical formulations of it: a negative one, 'what is hateful to you, do not do to your neighbour'; and a positive one, 'whatever you wish that men would do to you, do so to them'.[14] The negative formulation is most naturally read as intending to forestall the ever-present danger of the agent *taking advantage* of the power he has, and *imposing* his will on the recipient of his action.[15] The positive formulation is most naturally read as requiring something more: as a way of living in which one gives out to other people what one would like to receive from them. What is important is that both require the agent to cast herself into the role of the *patient* of her action, in order to appraise what she is doing.

---

[12] See Long & Sedley 1987, p.349

[13] My reading of cosmopolitanism is indebted to Martha Nussbaum: see especially Nussbaum (1997).

[14] These formulations are put forward by Rabbi Hillel (*Shabbath* 31a) and Matthew (7.12) respectively. For a good discussion of both, see Wattles (1996).

[15] Ricoeur (1992, p.219) underscores this point 'The most remarkable thing, however, in the formulation of this rule is that the reciprocity demanded stands out against the background of the presupposition of an initial dissymmetry between the protagonists of the action – a dissymmetry that places one in the position of agent and the other in that of patient. ... it is upon this dissymmetry that all the maleficent offshoots of interaction, beginning with influence and culminating in murder, will be grafted.'.

There are two important movements within the account of reciprocity; the first attempts to go beyond the golden rule, and demonstrate that a more carefully constructed principle of reciprocity can be shown to hold a priori; the second stays within the ambit of the golden rule, and seeks to clarify the *spirit* of the reciprocity that is required by it.[16]

In the first case, the golden rule is seen as a defective (even if intuitively appealing) statement of a more formal principle of reciprocity, which is supposed to hold a priori. Marcus Singer (1961, p.16) speaks for this whole approach when he says:

> In any of its traditional formulations this rule is not only imprecise, but if taken literally would be an abomination ... Such literal interpretations of the rule are undoubtedly misinterpretations of what is intended by it. But what this shows is that as it stands the rule is imprecise and needs qualification. It neither says what it means nor means what it says.

Sidgwick (1962, p.380) thought that the golden rule when properly stated was self-evident, and must take the following form:

> it cannot be right for A to treat B in a manner in which it would be wrong for B to treat A, merely on the ground that they are two different individuals, and without there being any difference between the natures or circumstances of the two which can be stated as a reasonable ground for difference of treatment.[17]

This seems to be the best place to slot Kant's formula of universal law: 'Act only according to that maxim whereby you can at the same time will that it should become a universal law'. This carries the formalist reworking of the golden rule to its logical conclusion. Not only does it exclude all reference to any contingent matters such as what the agent happens to

---

[16] There is a parallel here to the two ways of expressing the idea of cosmopolitanism: in both cases we have two different approaches to an abstract idea; one which attempts to show its objective necessity, and one which attempts to bring it closer to the agent's own concerns.

[17] Cf also Samuel Clarke's 'principle of equity': 'Whatever I judge reasonable or unreasonable for another to do for me; that, by the same judgment, I declare reasonable or unreasonable, that I in the like case should do for him.' (Clarke 1705; quoted in Marcus Singer 1961, p.16).

want, but it is also an extension by generalization of it: where the golden rule imagines that there is a situation with only two people, A and B, and attempts to equalize the relationship between agent and patient in this relationship, the categorical imperative attempts to equalize the relationship between the agent and *all other human beings.*[18]

In the second approach, the true intent of reciprocity is uncovered through reflecting on the golden rule. In doing so, it tackles head on the following problem (which is left untouched by even the most accurately stated formalization of the golden rule): the rule merely requires *consistency.* While it might require that I not harm others, it seems to say little about my *benefiting* them if they are in need (as Kant put it, 'many a man would gladly consent that others should not benefit him, if only he might be excused from benefiting them' (*G*, p.430n). This response to reciprocity attempts to make clear that such a reading of it is contrary to its spirit. It aims to take the underlying idea of reciprocity, and infuse it with a *generosity,* so that treating others as one would like to be treated then *contains within it* the idea of helping others, and not just refraining from harming them.

The Sermon on the Plain, which introduces the command to love one's enemies, provides a good example of this approach to reciprocity. I shall not attempt to engage with the vast amount of literature and commentary that this passage has engendered; nor shall I comment on any *religious* meaning that the passage may be supposed to have. I shall merely follow Paul Ricoeur's reading of how the rhetoric of this passage is supposed to impact on the idea of reciprocity and of morality in general. The passage in question runs as follows:

---

[18] Interpreted in *this* way, reciprocity is closely linked to cosmopolitanism (and, in the process, gives a new way of understanding cosmopolitanism): for we can get to the idea of cosmopolitanism by laying down as a rule of reciprocity that each must act on purposes that are capable of being universal laws. (Note, however, that Kant himself is quite dismissive of the golden rule in the *Groundwork*; see his note at (*G,* p.430.))

> But I say to you that hear, Love your enemies, do good to those who hate you, bless those who curse you, pray for those who abuse you. To him who strikes you on the cheek, offer the other also; and from him who takes away your cloak do not withhold your coat as well. Give to everyone who begs from you; and of him who takes away your goods, do not ask for them again. And as you wish that men would do to you, do so to them. If you love those who love you, what credit is that to you? For even sinners love those who love them. And if you lend to those from whom you hope to receive, what credit is that to you? For even sinners lend to sinners, to receive as much again. But love your enemies, and do good, and lend, expecting nothing in return; and your reward will be great, and you will be sons of the Most High (Luke, 6.27-35)

For Ricoeur, the command to love one's enemies is deliberately paradoxical. It is deployed strategically as part of an attempt to undermine the mean-minded idea of reciprocity, in which practice of the golden rule is merely a veiled form of self-interest, where one 'does unto others...' *in order to* further one's own interests.[19] The idea of *unilateral love* for enemies is deployed to stretch the underlying *bilateral* notion of reciprocity almost to breaking point, *so that* once this idea has been released, the underlying idea of reciprocity has been permanently enlarged.

In Ricoeur's words, in disorienting us with the paradox of love for enemies, it 'rescu[es] true reciprocity from its caricature' – that is, from 'the merchant's economy of exchange and its logic of equivalence', by the 'substitution of the new motive of generosity for the ancient motive of self-interest'. (1992, p. 395)

As before with the Hierocles fragment, a key factor in morality turns out to be the agent rethinking her own understanding of herself, *in order that* the pursuit of universalistic justice does not seem to her as something *aberrant,* but rather as something that flows naturally from herself.

The dignity of persons

---

[19] As the Romans said, in sacrificing to their gods: *do ut des.* ('I give *so that* you may give.')

The third way that morality attempts to render more concrete what it is to give each person what is due to them is through the claim that human beings as such have a *dignity*. In the tradition, this has been taken to entail the following three claims.

1. Human beings have a worth that is categorially above anything non-human

2. Human beings are *equal in worth to one another* (though, of course, each is categorially more valuable than anything that is not a human being).

3. Human dignity must be respected at all times.

Kant is the exponent par excellence of this conception of human dignity:

> Man as a person, i.e. as the subject of a morally-practical reason, is exalted above all price. ... [H]e possesses a dignity (an absolute inner worth) whereby he exacts the respect of all other rational beings in the world, can measure himself against each member of his species, and can esteem himself on a footing of equality with them. (*MM*, p.435)

This conception of human dignity obviously leads back to cosmopolitanism; if each human being has this absolute inner worth, and it is the same in each person, then it is easy to see how it could give another ground for cosmopolitanism as follows: the scope of justice should be universal because all human beings have dignity. This account of human dignity leads to a new formulation of the idea of acting well: acting well consists, above all, in *respecting* the dignity of each human being in the way that one acts. Hence the formula of the categorical imperative usually referred to as the formula of humanity:

> Act in such a way that you treat humanity, whether in your own person or in the person of another, always at the same time as an end and never simply as a means.[20] (*G,* p.429)

---

[20] See also (*G,* p.428): '[R]ational beings are called persons inasmuch as their nature already

It is from this strand that we get an absolute prohibition on forms of treatment designed to degrade, to destroy humanity in a person, such as torture:

> To show contempt (*contemnere*) for others, i.e., to refuse them the respect which is owed man in general, is in all cases contrary to duty; they are, after all, men. ... So it is an outrage to inflict punishments that dishonour humanity itself (such as drawing and quartering, letting someone be torn to pieces by dogs, cutting off noses and ears). Such punishments are not only more grievous to one who loves honour (who claims the respect of others, as everyone must) than is the loss of life and possessions; they also make a spectator blush to belong to a race that can be treated that way.[21] (*MM,* p.463)

There is much that is unmistakeably attractive in this conception of human dignity. For it solves brilliantly the central problem of moral motivation, by explaining the *point* of morality in terms of the incomparable worth of each and every human being. So it is with a heavy heart that we conclude in part two that this Kantian conception of human dignity is untenable, and in part three attempt to construct something to replace it out of the resources available to us.

1.3 *On the varieties of morality*

Morality (as a tradition) also contains within it deep disagreements.[22] I shall consider two: first, the dispute about the status of rules in morality; second, the more wide-ranging and intractable dispute about the relationship between morality and human nature.

---

marks them out as ends in themselves, i.e. as something which is not to be used merely as means and hence there is imposed a limit on the arbitrary use of such beings, which are thus objects of respect.' Note how this dovetails with the requirement of morality that it be a *praxis,* not a poiesis.

[21] This underlying idea has of course now been given a legal sanction in the United Nations Declaration on Human Rights: Article 5 states that 'No one shall be subjected to torture or to cruel, inhuman or degrading treatment or punishment', and incorporated into British Law via the 1998 European Convention on Human Rights and Human Rights Act, article 3.

[22] Remember MacIntyre's definition, quoted above, of a tradition as 'an historically extended, socially embodied argument, and an argument precisely in part about the goods which constitute that tradition.'

i) Duty, rules and moral life

All positions that fall within the tradition of morality must make a place for universal moral rules and for the motive of duty. No ethical position could claim to be a part of the tradition of morality if it refused to admit that there must be a coercive, rule based side to morality whose aim is to *ensure that the dignity of human beings is respected at all times.* It cannot, for example, merely leave it up to the individual's own choice whether or not he chooses to respect the dignity of others. The reason for this is simple: the tradition of morality believes that respect is due whether or not someone happens to want to give it at a particular time, and hence it must take steps to ensure that there are moral rules in place that can be enforced and which serve to protect the dignity of each.

The motive of duty (acting out of respect for moral rules) is the internal analogue of universal moral rules: its aim is to create a psychological structure such that the individual is able to *coerce* himself into acting morally if need be. Such self-coercion is necessary if we are to do what is morally required in a situation in which we find ourselves shrinking from it: for example, when we are faced with a person whom we greatly dislike, but must nonetheless treat with the respect due to them as a human being.

However, there is a deep dispute within morality as to whether moral rules and accountability to them constitute the *essence* of morality, or if they are just its baseline, below which we must not fall, but beyond which more spontaneous and joyful expressions of the spirit of morality are possible.

Kant is the leading spokesman for the part of the tradition that takes moral rules and duty to form the essence of morality: on his view it is dangerous arrogance to suppose that we can raise ourselves above the need to hold ourselves continually accountable to moral rules. Moreover, if we insert

any motive other than doing our duty for its own sake into our doing what is morally required, we deprive our action of all moral worth:

> It is a very beautiful thing to do good to human beings because of love and a sympathetic good will, or to do justice because of a love of order. But this is not yet the genuine moral maxim of conduct, the maxim befitting our position among rational beings as men, when we presume, like volunteers, to flout with proud conceit the thought of duty and, as independent of command, merely to will of our own good pleasure to do something to which we think we need no command. We stand under a *discipline* of reason, and in all our maxims we must not forget our subjection to it, or withdraw anything from it, or by an egotistical illusion detract from the authority of the law (even though it is one given by our own reason), so that we could place the motive of our will (even though it is in accordance with the law) elsewhere than in the law itself and in respect for it. (*C2, p.82*)

This approach has a knock-on effect on theories of moral judgment. It places the primary locus of moral judgment in the rules themselves: moral judgment consists in bringing the case under a rule, 'regarding its concrete features as ethically salient insofar as they are instances of the universal.' (Nussbaum, 1986, p.300) It tends to talk not of rules, but of moral *laws,* which it will take to hold universally and to be as implacable and as unconditional as all laws. We find this approach to moral judgement is marked in the Cicero's definition of natural law, in the formalistic reworkings of the golden rule, and above all in Kant's account of morality.

On the other hand, we have positions which see moral rules and acting from the motive of duty as only part of morality. On this view, morality appears under the light of moral rules and of duty only to those who are not truly good. For those who have made more progress towards goodness, morality expresses itself more typically in a spontaneous generosity and giving: duty and moral rules are still there in the background, and may

come into play when the more spontaneous goodness is for some reason or other not currently firing, but they do not form the soul of moral life.[23]

This conception of rules leads to its own account of moral judgment. It tends to play down the importance of moral rules: rather than seeing moral rules as unconditionally valid in their own right, it will suggest that we do better to think of them as imperfect attempts to spell out in a perspicuous and easily understood way the meaning of the morally basic notions of reciprocity, cosmopolitanism and human dignity. Moral rules, on this conception are *summaries* of good concrete judgments; their force as principles comes from the wisdom of these judgments not from anything intrinsic in the rule as such. Moral rules will be subject to revision in the light of further insight; and it may even allow exceptions to be made to these moral rules *in the name of morality.* In this camp would be Paul Ricoeur's reworking of reciprocity, and (I think) Hierocles' understanding of cosmopolitanism in terms of the concentric circles.

ii) The relationship between morality and human nature

This terrain is far more difficult to map; the disagreements within morality about its relationship to human nature are deep and tangled. But our purposes do not require us to produce a complete map of all the variations that are possible on this topic within morality; it will be sufficient for this introductory chapter if we roughly indicate some of the terrain that such a map would provide the key to.

a) There is a dispute about whether morality, as such, even stands in need of an account of human nature; as we shall see in the next chapter, on Kant's account of it, it does not ('when applied to man, it does not in the

---

[23] I take it that this is the view expressed for example by Iris Murdoch, and it is close to the view I shall argue for in chapter eight. "A realistic view of morality cannot dispense with the idea [of duty]; duty is for most people the most obvious form of moral experience: Kant's starting point. One might say that morality divides between moral obligation and spiritual change. The good life becomes increasingly selfless through an increased awareness of, sensibility to, the world beyond the self. But meanwhile requirements and claims, which we still recognise abstractly and as it were externally, demand to be met." (Murdoch, 1992, p. 53)

least borrow from acquaintance with him (anthropology) but gives a priori laws to him as a rational being' *G*, p. 389). Alternatively (as I shall argue against Kant) knowledge of human beings could be most relevant to morality.

b) Next, there is a dispute about the relationship between moral virtue and normal human development; that is, is morality something that normally brought up human beings just grow into? We have positions here ranging from Aristotle-inspired accounts, which take inculcation into moral virtue to be normal and natural, to Augustine-inspired accounts, which take it that human beings are naturally bad, and must be kept in line by the firm force of moral law.

c) Next there is a similar dispute that occurs, not in relation to human development, but in relation to the general goal of human life; that is, how is morality related to human flourishing? Here we have positions ranging from the highly optimistic, such as Aquinas', which see morality and the best life for a human being as of one piece to positions, which state that there is no necessary relationship at all, but duty is still duty (see e.g. Prichard 1912).

d) Last, there is a dispute as to why human beings deviate from what is morally required; or in more traditional language, why does moral evil occur? There are many answers within the tradition, and they do not form a natural continuum. Important answers would be: human beings are naturally bad; our bodily nature tempts us away from the pure path of duty; human beings find it easier to divert their attention away than to respond to the call of morality.

To sum up: morality has both a common core – consisting in the interwoven ideas of cosmopolitanism, reciprocity and human dignity, and an internal diversity – based both on different understandings of the role of

rules and on manifold different understandings of the relationship between morality and human nature.

I conclude this chapter with a few brief words on the end at which morality aims. What we can say about the end of morality in general is slight, but yet important; it consists in the realization of the ideals of cosmopolitanism, reciprocity and respect for human dignity in a community, and it is in acting in such a way as to make this community possible that morality consists. But any attempt to further specify this end will be subject to variation within the tradition; there will be disputes related to the variations we have just mentioned as to what would *constitute* a realization of this end, what practical steps we should take to move towards it and the degree to which it is possible for a human community to approximate to it.

Chapter Two
Against Kant, for the sake of Morality

2.1 Introduction

This chapter is devoted to Kant's account of morality. Kant bears a complex relationship to the project of this thesis: I have no quarrel with the usual evaluation that Kant is the greatest systematizer that the tradition of morality has known, but at the same time the conception of morality for which I shall argue in part three is as far as it is possible to get from Kant's whilst still remaining within the tradition.

Kant's claim to greatness as a systematizer lies in the way he takes the fundamental ideas of morality (cosmopolitanism, reciprocity and human dignity) and interlinks them in a new and rigorous way through his different formulations of the categorical imperative. We saw in the last chapter how the formula of universal law could be seen as an extension and generalization of the reciprocity of the golden rule; how the formula of humanity is a brilliant exposition of what it is to respect the dignity of other human beings; and the end at which morality aims is well captured in the formula of the kingdom of ends.

However, Kant combines these insights into the fundamental ideas of morality with certain other ideas that, whilst unnecessary to morality as such, make his account of morality both self-undermining (and hence impossible) and offensive to many of the ethical insights morality should be trying to safeguard. As we shall see, there are several such ideas. What draws all these ideas together is their common desire to make morality as *pure* as possible, both by sharply demarcating morality from every other

field of human endeavour and by purging morality of all elements of contingency. Thus for example, moral rules are said to hold a priori, and to have no element of empirical determination in them; respect for the moral law is posited as a special form of motivation that is not contaminated by anything sensuous or emotional; acting on any motivation other than respect for the moral law is denigrated as not possessing true moral worth.

This chapter critiques this emphasis on purity in Kant's conception of morality. It falls into three parts. Section 2.2 argues that there is no reason to follow Kant's purifying assumptions about the possibility of morality: morality is quite possible without these assumptions.

Section 2.3 argues that morality is in fact *impossible* if we take Kant's presuppositions about the moral law seriously. Section 2.4 argues that there are good reasons for thinking that the view of human life that comes out of it gets so much wrong that Kant's account of morality is inappropriate for the general regulation of human behaviour.

2.2 *Kant's presuppositions about morality*

Kant thinks that if there is to be such a thing as morality (and hence moral obligation) at all, it is only on the basis of an account that makes morality totally pure and free from all contingency, as his account does. What is essential to Kant's account of morality is his insistence that the obligatory character of morality can be understood only if we recognize that moral rules must be valid a priori for all rational beings and we, as rational beings, are so constructed that we can and should act on the basis of a priori moral rules alone.

In the preface to the *Groundwork*, Kant purports to find this requirement for purity in 'the common idea of duty and of moral laws'. He continues as follows:

Everyone must admit that if a law is to be morally valid, i.e., is to be valid as a ground of obligation, then it must carry with it absolute necessity. He must admit that the command, "Thou shalt not lie," does not hold only for men, as if other rational beings had no need to abide by it, and so with all other moral laws properly so called. And he must concede that the ground of obligation here must therefore be sought not in the nature of man nor in the circumstances of the world in which man is placed, but must be sought a priori solely in the concepts of pure reason; he must grant that every other precept which is founded on principles of mere experience – even a precept that may in certain respects be universal – insofar as it rests in the least on empirical grounds – perhaps only in its motive – can indeed be called a practical rule, but never a moral law. (*G,* p.389)

Kant has a Janus-faced argument for his conception of morality in the *Critique of Practical Reason.* I call it Janus-faced because it seeks simultaneously to explain how an a priori moral law of the sort Kant thinks that there is is possible, and to show that moral obligation would not be possible on any other basis. But before we can evaluate this argument, we must introduce some Kantian technical terms.

2.2.1 Definitions of terms

Kant's argument requires us to distinguish *maxims, practical rules* and *imperatives.* All are *practical principles,* that is, 'propositions which contain a general determination of the will' (*C2,* p.18). Maxims are *subjective,* and describe the general conditions under which an agent is acting or will act (e.g. 'I will steal whenever I think I can get away with it, and the stealing would not harm a friend').[24] Practical rules are *objective,* and are expressive of rationality. We, as finite (embodied) beings, experience practical rules as *imperatives,* that is, as instructions telling us how we *ought to* act.[25] In Kant's words, an imperative is a 'rule characterized by an 'ought' which expresses the objective necessitation of

---

[24] Cf the definition of a maxim at *G,* p.421n.

[25] A holy being cannot be tempted to deviate from practical rules, and so automatically acts in accordance with the deliverances of morality. A holy being therefore does not experience the moral laws as accompanied by an 'ought', or feel *commanded* by morality.

the act and indicates that, if reason completely determined the will, the action would without exception take place according to the rule.'[26] (*C2,* p.20)

It is obvious that any account of morality must hold that moral rules are imperatives of some sort. For, by its very nature, morality says that there are standards implicit in cosmopolitanism, reciprocity and the human dignity that human beings *ought* to live up to. Moreover, it is equally obvious that we are not holy beings for whom acting in accordance with these standards 'inheres by an *inner* necessity,' (*MM,* p.222) so morality as such must go along with the idea that moral rules are (for us) imperatives.

Kant subdivides the class of imperatives into two: *hypothetical* and *categorical.* The way that Kant draws the distinction in the *Critique of Practical Reason* is somewhat elusive, so I shall draw also on the *Metaphysics of Morals.* The *Critique of Practical Reason* account goes as follows:

> Imperatives determine either the conditions of causality of a rational being as an efficient cause only in respect to its effect and its sufficiency to bring this about, or they determine only the will, whether it be adequate to the effect or not. In the former case, imperatives would be hypothetical and would contain only precepts of skill; in the

---

[26] The counter-factual reference to what would happen *were* reason to solely determine the agent's will is problematic, and reveals that Kant has not yet fully thought through the *Wille/Willkür* distinction. In *MM,* Kant distinguishes between the *will,* considered as practical reason (Wille) from *choice,* considered as lack of determination by any sensible determining grounds (*Willkür*). The point of this distinction is to explain how we can be responsible for our actions when we act heteronomously – something that is somewhat opaque on the account given in *C2* as it stands.

In this reconstruction I factor the *Wille/Willkür* distinction into my account, and interpret Kant's writing in terms of it whenever it is needed by what he wants to say, even if he is not explicit in adverting to the distinction. In *MM,* Kant defines an imperative in a way that is compatible with the *Wille/Willkür* distinction: 'An imperative is a practical rule by which an action in itself contingent is *made* necessary. An imperative differs from a practical law in that a law indeed represents an action as necessary but takes account of whether this action already inheres by an *inner* necessity in the acting subject (as in a holy being) or whether it is contingent (as in man): for where the former is the case there is no imperative. Hence an imperative is a rule the representation of which *makes* necessary an action that is subjectively contingent and thus represents the subject as one that must be *constrained* (necessitated) to conform with the rule.' (*MM,* p. 222)

latter they would be categorical and would alone be practical laws. (*C2,* p.20)

What he says elsewhere allows us to supplement this account as follows: hypothetical imperatives are all means-end – that is, they presuppose some end or goal and *confer necessity* on the means to be followed in pursuing this end. A categorical imperative, on the other hand, 'represents an action as objectively necessary and makes it necessary not indirectly, through the representation of some *end* that can be attained by the action, but through the mere representation of this action itself (its form) and hence directly.' (*MM,* p.222)

On the most literal reading of the *Metaphysics of Morals* account, the distinction would appear to be that hypothetical imperatives presuppose some end, and categorical imperatives are completely formal and hence abstract from all ends. But this cannot be right. Kant cannot simultaneously have claimed that categorical imperatives abstract from all ends *and* that we can formulate the categorical imperative in terms of treating humanity whether in one's own person or in the person of another as an end and never simply as a means.

But, if we turn to the *Groundwork,* we discover that what it is for a principle to be *formal* is subtly different from this. A formal principle does not have to abstract from *all* ends, just *subjective* ends:

> Practical principles are formal when they abstract from all subjective ends; they are material, however, when they are founded upon subjective ends, and hence upon certain incentives. (*G,* p.427)

Thus we can read the formula of humanity as telling us to take humanity as an *objective* end:

> If then there is to be a supreme practical principle and, as far as the human will is concerned, a categorical imperative, then it must be such that from the conception of what is necessarily an end for everyone because this end is an end

in itself it constitutes an objective principle of the will and can hence serve as a practical law. (*G,* p.428)

I discuss below how objective ends are related to the categorical imperative; for the moment I shall take the idea of objective ends as read.

2.2.2 *Kant's argument for the a prioricity of moral laws*

Kant can now avail himself of the following argument in order to draw the conclusion that moral obligation must require a categorical imperative:

1) Morality must present itself in imperatives.

2) There are two types of imperatives: hypothetical and categorical.

3) Hypothetical imperatives are not adequate to moral obligation.

4) *Conclusion:* Therefore moral obligation requires a categorical imperative.[27]

There are problems with both 2) and 3). I shall discuss them in turn.

2) The argument requires that the division between categorical and hypothetical imperatives, each defined as Kant defines them, is exhaustive of all possible imperatives.

If we can imagine an imperative that would be sufficient for moral obligation, but which did not conform to the conditions that Kant places on a categorical imperative, then this argument will be shown to be unsound, for it will not follow from the fact that hypothetical imperatives are insufficient for moral obligation that morality requires a categorical imperative, defined as Kant defines it.

There are two obvious sorts of imperative that would be sufficient for moral obligation, but which differ from the categorical imperative as Kant conceives it.

---

[27] Kant does not, of course, set the argument out quite so obviously.

First, we could conceive of a *substantive* a priori principle of practical reason, of the sort that moral realists believe to hold, for example that pain is objectively bad.

Second, we could conceive of the moral point of view as being given a priori and capable of generating inescapable moral rules that take priority over all other practical reasons, but conceive of the moral rules themselves as constructed a posteriori and in relation to our actual human situation.

Whilst moral imperatives of either of these sorts would not be categorical imperatives under Kant's definition, they would certainly be *inescapable* and could claim priority over all other practical reasons, and so could provide a *law* (have unconditional validity) for human beings.[28]

3) Much more significant, however, is the assumption behind this argument, namely that if morality is to be possible at all, moral imperatives must have the force of *law* (have unconditional validity). Kant attempts to substantiate this claim by arguing in the following way that hypothetical imperatives could not be sufficient for moral obligation:

i) We could not expect the subjective ends presupposed by hypothetical imperatives to hold universally.

ii) Even if the end held universally and was inescapable for us, a hypothetical imperative still could not be sufficient for moral obligation, because the end would be merely *physically necessary* and not *rationally necessary.*

iii) The end in a hypothetical imperative would be contingent relative to other ends that the agent had, and hence it could never provide a *law* (that

---

[28] Another sort of imperative that might plausibly be thought to be sufficient for moral obligation would be one resembling the hypothetical imperative, but which posits an *objective end,* and explains moral obligation as following from the necessity of this end. Kant has a separate argument against such a position, that I shall allude to later in the discussion of the relationship between the categorical imperative and objective ends.

is, have unconditional validity), which would be necessary for moral obligation.

There is a whiff of circularity here: if you presuppose that moral obligation is possible *only* on the basis of unconditionally binding moral laws, then it follows that you cannot get moral obligation out of hypothetical imperatives. But why should one suppose such strict conditions attend the possibility of moral obligation? I believe that there is no good reason, particularly given that (as we shall see) moral obligation is in fact impossible on the basis of the account Kant gives of it. Rather, I am inclined to see in Kant a certain (understandable) lack of confidence in the worthwhileness of the moral life on his conception of it, and a correlative over-compensation by insisting over-enthusiastically on the unconditional binding force of moral obligation.[29]

For, if we could find a deep desire that is so closely connected to our humanity that we think that we would not continue to be the sort of beings we are without it, and show that this desire leads inexorably to accepting the validity of moral obligations, then it is hard to see why such an account of moral obligation would not be sufficient for morality. As we shall see in part two, Christine Korsgaard and Charles Taylor develop this thought in two different ways. Korsgaard aims to show that valuing one's self and one's identity at all presupposes valuing one's humanity (one's ability to value), and that this chain of reasoning should lead one to value the humanity of all human beings. Charles Taylor suggests that our deep need to make sense of our lives points us on a similar quest, whose end goal is the valuing of the dignity of all human beings.

More radically, and as I shall argue in chapter eight, we could adopt an even more hypothetical reading of moral obligation: rather than attempting to hitch moral obligation to something that is contingent in human beings but yet constitutive of human selfhood, we could explain moral obligation

---

[29] I spell this thought out in more detail in 2.4.

as much more optional. If we take it that the purpose of moral rules is the general regulation of human life in such a way that the dignity of each is protected, and that the concepts of moral obligation and of duty are to be explained in terms of this goal, then we do not *need* to conceive of moral obligations as categorical imperatives. We could develop a system of moral obligations in which those obligations are recognised as hypothetical, dependent on a desire to respect the dignity of others and the desire to guarantee that the dignity of others is respected by the promulgation of general rules.

2.3 *The impossibility of moral obligation on Kant's account of morality*

We have seen that the strict bisection of imperatives into hypothetical and categorical, and the claim that moral obligation requires a categorical imperative, are in no way necessitated by the project of explaining the *possibility* of moral obligation. However, this strict bisection into hypothetical and categorical imperatives plays a basic and structural role in Kant's entire account of morality. To see how, let us take up the argument of the *Critique of Practical Reason* again, this time from the point at which Kant introduces his first theorem:

> All practical principles which presuppose an object (material) of the faculty of desire as the determining ground of the will are without exception empirical and can hand down no practical laws. (*C2,* p.21)

This theorem decomposes into two claims:

i) All practical principles which presuppose an object of the faculty of desire as the determining ground of the will are empirical.

ii) No empirical practical principle can be a practical law.

Both of these claims, on closer examination, turn out to be merely restatements of the fundamental distinction between hypothetical and

categorical imperatives. Thus, the argument he gives in support of i) is as follows:

> [T]he determining ground of choice consists in the conception of an object and its relation to the subject, whereby the faculty of desire is determined to seek its realization. Such a relation to the subject is called pleasure in the reality of an object, and must be presupposed as the condition of the possibility of the determination of choice. But we cannot know, a priori, from the idea of any object, whatever the nature of this idea, whether it will be associated with pleasure or displeasure or will be merely indifferent. (*C2,* p.21)

This merely states what I have already adduced as Kant's first argument that morality cannot be based on categorical imperatives; except here that claim is not that we cannot expect a desire to hold universally, but that we cannot tell a priori if an object will provoke desire. Similarly, Kant's argument for ii) is a mixture of the second and third arguments why morality cannot be grounded on hypothetical imperatives:

> A principle which is based only on the subjective susceptibility to a pleasure or displeasure ... cannot function as a law even to the subject possessing this susceptibility, because it lacks objective necessity, which must be known a priori. (*C2*, p.21-22)

I labour this point, because it seems to me to be of fundamental importance to understanding the deep structure of Kant's account of morality. As we progress through the *Critique of Practical Reason* we encounter a series of oppositions: empirical vs. a priori; hypothetical imperatives vs. categorical imperatives; empirical principles vs. practical laws; material vs. formal; determination by desire vs. lack of determination by desire. What is important is that all those on the left hand side of these oppositions are continually explicated in terms of one another, and likewise for the ones on the right hand side. It is somewhat misleading, then, to think of these as a *series* of oppositions; rather, there appears to be one fundamental opposition, of which these individual ones are partial explications.

This point is most important for the interpretation of Kant's account of morality; for it allows us to see that, given that these oppositions continually refer back to one another, we have far less room to manoeuvre than the sympathetic interpreter of Kant might like. There is no prospect of bringing an a posteriori account of moral imperatives such as the one that I canvassed as an alternative to the categorical imperative into Kant's account of morality; nor an account of desire and emotion that allows that they have a share in rationality. For to do so would be to unravel the oppositions that are constitutive of Kant's account of morality.

There is one final opposition that takes all those we have introduced so far, and extends them. This is the opposition between 'self-love or one's happiness' and morality. Kant's second theorem connects the 'self-love or one's happiness' vs. morality opposition to all the others that we have so far encountered:

> All material principles are, as such, of one and the same kind and belong under the general principle of self-love or one's happiness. (*C2*, p.22)

This claim requires some interpretation. Kant explicates the notion of happiness as follows:

> [H]appiness is a rational being's consciousness of the agreeableness of life which without interruption accompanies his whole existence, and to make this the supreme ground for the determination of choice constitutes the principle of self-love. (*C2*, p.22)

The important thing to note here is that the distinction between morality and self-love is supposed to lie at the level of determination of the will; the principles of morality and self-love are two fundamental *orientations* toward the moral law with which one might live one's life.[30] What makes material principles fall under the general principle of self-love is that the *determining ground of the will* is inclination, not a priori principles; we

---

[30] We discover in the *Religion* that, according to Kant, these two fundamental orientations correspond to the difference between *good* and *evil* respectively. (*R*, p.31f.)

thus loop back to the discussion of formal vs. empirical principles, and from thence to the initial distinction between categorical and hypothetical imperatives.

Kant now moves to derive the principle of morality. From the definition of a practical principle (quoted above) we know that practical principles as such 'contain a general determination of the will'. We have just seen that all material principles fall under the principle of self-love. Given that the moral principle is defined in *opposition* to the principle of self-love, it cannot be a material principle. Putting these two thoughts together Kant arrives at his third theorem:

> Theorem III: If a rational being can think of his maxims as practical universal laws, he can do so only by considering them as principles which contain the determining grounds of the will because of their form and not because of their matter. (*C2*, p.26-7)

A more expansive working over of this point would be as follows: all material principles (that is, those that have their determining ground in something empirical, i.e. desire) have been shown not to furnish any practical laws. Hence if there is to be a practical law, the determining ground of the will cannot be material, and 'If all material of a law, i.e. every object of the will considered as a ground of its determination, is taken away from it, nothing remains except the mere form of giving universal law.' (*C2*, p.27)

From here, Kant moves swiftly to deduce what a practical law must be like:

> Since the material of the practical law, i.e. an object of the maxim, cannot be given except empirically, and since a free will must be independent of all empirical conditions (i.e. those belonging to the world of sense) and yet be determinable, a free will must find its ground of determination in the law, but independently of the material of the law. But besides the latter there is nothing in a law except the legislative form. Therefore, the legislative form,

insofar as it is contained in the maxim, is the only thing
which can constitute a determining ground of the free will.[31]
(*C2,* p.29)

It follows from this that a practical law must contain only its fitness to be
legislated as universal law. With this we come back full circle to the
discussion of the distinction between hypothetical and categorical
imperatives. Now we discover that the categorical imperative, as what a
practical law must appear to us as, must take the following form:

So act that the maxim of your will could always hold at the
same time as the principle giving universal law. (*C2,* p.30).

2.3.1 *Objective ends and the categorical imperative*

It is time to return to a discussion I postponed earlier. We must elucidate
how objective ends are related to the above formulation of the categorical
imperative, which Kant describes as the 'Fundamental Law of Pure
Practical Reason' (*C2,* p.30). There are three basic relations that objective
ends could have to the above formulation of the categorical imperative:
they could either be *prior* to it, *simultaneous* with it, or *posterior* to it.

Objective ends would be *prior* to the categorical imperative if the source of
the obligation of the objective ends were more fundamental than that of the
categorical imperative, so that the obligation of the categorical imperative
was regarded as following from these ends.

Objective ends would be *simultaneous* with the categorical imperative if
the source of obligation of the categorical imperative was the same as that
of the objective ends. This would be the case either if objective ends and
the categorical imperative logically entailed one another or if taking up
objective ends were a *necessary effect* of the categorical imperative on a
finite being.

---

[31] I pass over the preceding section (§5) in which Kant attempts to show that a will that can be
determined by the mere legislative form of maxims must have its determining ground 'distinct from
all determining grounds of nature', which 'independence is called *freedom* in the strictest, i.e.,
transcendental sense'. (*C2,* p.29)

Objective ends would be *posterior* to the categorical imperative if the source of obligation of the categorical imperative was prior to that of the objective ends and did not in any way entail the existence of objective ends. On this view, objective ends would be something over and above what we are committed to by the categorical imperative.

Kant explicitly argues that objective ends cannot be prior to the categorical imperative. See for example how he lambasts perfectionism:

> Only if ends are already given can the concept of perfection in relation to them (either internal perfection in ourselves or external perfection of God) be the determining ground of the will. An end, however, as an object which precedes and contains the ground of determination of the will by a practical rule – that is, an end as the material of the will – is, if taken as a determining ground of the will, only empirical; it could thus serve for the Epicurean principle in the happiness theory but never as a pure rational principle of ethics and duty. (*C2,* p.41)

Indeed, the whole idea of an ethics of autonomy is based on a rejection of the priority of objective ends to the categorical imperative.

This leaves the other two options. In order to see what is at stake in our answer to the question, it may be helpful to consider it from a different perspective: what is the relationship between the formulation of the categorical imperative in terms of universal law and that in terms of treating humanity as an end in itself?

We can see the remaining two options (priority and simultaneity) as attempts to explain this relationship. The first form of the simultaneity option states that the two formulations obviously and objectively state the same thing and hence have the same consequences for action. Kant suggests this in the *Groundwork* where he baldly states that:

> The aforementioned ways of representing the principle of morality are at bottom only so many formulas of the same law: one of them by itself contains a combination of the

other two. Nevertheless there is a difference in them, which is subjectively rather than objectively practical ... But one does better if in moral judgment he follows the rigorous method and takes as his basis the universal formula of the categorical imperative. (*G,* p.436-7)

The difficulty with this is that it merely asserts its conclusion rather than arguing for it or explaining *why* it is the case: it is far from obvious that the formula of universal law and the formula of humanity are the same at the objective level.

The second form of the simultaneity option tries harder to explain *why* the formula of universal law and the formula of humanity coincide. Its basic thesis is this: the formula of universal law is the categorical imperative considered *objectively*; the formula of humanity is the categorical imperative considered *subjectively.* When considered *objectively,* the categorical imperative abstracts from all ends; when considered *subjectively* (that is, in terms of the effect it has on finite rational beings) it prescribes humanity as an end in itself, which is then called an *objective end.* The formula of universal law and the formula of humanity thus coincide because they are, so to speak, the inside and outside of the same thing.

In the priority option, where objective ends are taken to be posterior to the categorical imperative, the formula of humanity will be taken to be posterior to the formula of universal law and not derivable from it. The formula of universal law will assume a lexical priority in moral deliberation. The problem with this approach is that it seems impossible to square with the argument of the *Critique of Practical Reason.* Kant's argument is dedicated to uncovering what a *practical law* must be like; the whole structure of the argument assumes that there can only be *one* practical law. Further, there is no attempt to give the formula of humanity a separate derivation. Where we *do* have an attempt at a separate derivation, this either sits side by side with the claim that the various formulations of

the categorical imperative are objectively identical (as at *G,* p.428) or does not seem to be compatible with the argument of the *Critique of Practical Reason* (as at *MM,* p.385). Now, while it is true that we could attempt to reconstruct a Kantian account of morality *on the basis* of these passages, to do so would be to ride roughshod over the way that Kant wants us to construe his account of morality: the essence of his account is to be found in its foundations, and not, for example, in the *applications* of the categorical imperative.

What we see here is that there is a tension between the account Kant wishes to adopt to ground moral obligation, and the more sophisticated account he sometimes wishes to adopt when actually spelling out moral obligations. We are reconstructing his view from the perspective of his attempted justification of the categorical imperative; so we must leave on one side attempts such as these which might trace the formula of humanity back to a different source of obligation.[32]

It follows that, for the purposes of the examination we are undertaking here, we must assume that there is a congruence between the results of correct application of the formula of universal law and the formula of humanity; hence it will follow from Kant's argument that *if* the formula of universal law proves to be defective, then so necessarily must the formula of humanity.

2.3.2 *The impossibility charge*

The impossibility charge states that i) Kant is committed, as an *essential and ineradicable* part of his account of morality, to certain claims and ii) it is impossible for these claims to be true.

I shall investigate only one of the claims of impossibility that could be made against Kant's account of morality – that against the fundamental

---

[32] In chapter four I give an exhaustive analysis of Christine Korsgaard's argument in the *Sources of Normativity,* which attempts reconstruct Kantian ethics through such a separation between the formula of universal law and the formula of humanity.

proposition Kant seeks to demonstrate in the *Critique of Practical Reason,* namely that pure reason can be practical. There are various facets of Kant's account that could be claimed to be impossible, such as the 'transcendental' freedom, or the postulates of practical reason. I am not unsympathetic to these charges; but prosecution of them would not take us any further toward our task of constructing a viable account of morality. I have therefore left them on one side, and have concentrated on a claim that is of immense importance for our broader project: that pure reason can be practical.

The claim that pure reason can be practical is highly compressed. Kant provides an informative gloss on it when he introduces the main question of the *Critique of Practical Reason* as follows:

> Is pure reason sufficient to determine the will, or is it only
> as empirically conditioned that it can do so? (*C2,* p.15)

Pure reason can be practical, then, only on the condition that it can determine the will without relying on any empirical conditions. 'Empirical conditions' are, first of all, desire. Second, they must include any facts known only through experience; that is, facts which cannot be shown to be necessary a priori.

Thus pure reason is sufficient to determine the will if and only if the will can be determined by a maxim that incorporates nothing beyond what can be given a priori by the nature of reason as such.

The practicality of pure practical reason requires the truth of three sub-claims: I call these the *moral precepts claim,* the *moral judgment claim* and the *moral motivation claim.*

i) *Moral precepts:* For the practicality of pure reason to be possible, we must be able to derive *a priori* a stock of moral precepts, which are unconditionally binding; for example, that one should never tell a lie.[33]

---

[33] Kant sometimes calls such precepts *moral laws* (see e.g. *G,* p.389 - quoted above); I prefer

Pure reason would fail to be practical if we could not derive an adequate stock of precepts from it. It would also fail to be practical if the precepts pure reason was able to derive were not those that morality *ought to uphold.*

ii) *Moral judgment:* The ability to derive a stock of valid moral precepts would not be sufficient for pure reason to be practical, however. For we would also need to i) grasp which precepts were relevant to a particular situation and ii) grasp what the correct thing to do was, *given* that a certain set of precepts were relevant to the situation. Now, obviously, if pure reason is to be practical, then the determination of choice by judgment must be able to be completely abstracted from anything empirical. Pure reason will fail to be practical if, for example, judgment leaves it *indeterminate* what is to be done, or if there are *irresolvable dilemmas* about what is to be done in a given situation, or if judgment requires *empirical determining grounds* of the will.

iii) *Moral motivation:* if pure reason is to be practical, then not only must we be able to derive a stock of moral precepts a priori, and moral judgment must be possible, but also such precepts and/or judgments must also be able to lead to action; pure reason *in separation from inclination* must be in itself sufficient to motivate an agent to do what is morally required.

To demonstrate the impossibility charge we do not need to go so far as the moral motivation claim; we shall see that the categorical imperative will break down both at the level of moral precepts and of judgments, and hence the claim that pure reason can be practical cannot be maintained whether or not Kant's account of moral motivation is correct. Our confrontation with Kant's account of moral motivation will occur in the critique of his account of human life in 2.4.

---

moral precepts, as it forestalls a possible confusion between 'middle principles' and the moral law itself.

2.3.4 *The categorical imperative and the practicality of pure reason*

We can detect two possibly incompatible approaches in Kant to the task of demonstrating that pure reason can *both* supply moral precepts *and* provide situational judgment. The first predominates in the *Groundwork*: the second in the *Metaphysics of Morals.*

The first approach collapses the problem of moral precepts and the problem of moral judgment into one another; it maintains that we can apply the categorical imperative in a situation, and find out directly what we are to do. On this view, the categorical imperative *simultaneously* provides moral precepts and moral judgment.

The second approach keeps the problem of moral precepts and the problem of moral judgments separate. It takes it that the task of moral philosophy is to derive valid moral precepts by the application of the categorical imperative. Moral judgment, on this view, cannot be codified, but instead relies on a certain *skill* which can be learnt only through practice and experience.

The first approach is exemplified in the *Groundwork;* here we hear that the categorical imperative is the *canon* by which we are to morally estimate our actions.[34] Now, if pure reason is to be practical, Kant must mean something quite specific by 'canon' here: he must mean that passing the test of the categorical imperative is *constitutive* of the moral permissibility of any action. The categorical imperative cannot be, for example, merely a good rule of thumb, or a test that is only applicable to certain classes of case. For in either of these cases we would need to suppose some standard of moral judgment *to which* the categorical imperative was accountable. *If* (*per impossible*) this further standard were given in pure reason then perhaps the situation might be salvageable. But, as we saw above, it is

---

[34] 'We must be able to will that a maxim of our action become a universal law; this is the canon for morally estimating any of our actions.' (*G*, p.424)

essential to Kant's argument that the categorical imperative is *the* fundamental law of pure practical reason.[35]

Given that there cannot be a standard behind the categorical imperative which could correct it, the categorical imperative must stand on its own as *the* canon for morally estimating actions. The problem, as we shall see, is that the categorical imperative, described in terms of the formula of universal law, cannot serve as a canon for morally estimating actions; for it simply fails to pick out all and only those maxims that it is morally permissible to act on. This is the case whichever of the different readings may be given of the formula of universal law ('Act only according to that maxim whereby you can at the same time will that it should become a universal law.')

Any maxim that generates a contradiction when the agent attempts to will it as a universal law is to be rejected; those that do not are morally permissible. Kant states that there are two ways in which a maxim can fail the test:

> Some actions are so constituted that their maxims cannot without contradiction even be thought as a universal law of nature, much less be willed as what should become one. In the case of others this internal impossibility is indeed not found, but there is still no possibility of willing that their maxim should be raised to the universality of a law of nature, because such a will would contradict itself. There is no difficulty in seeing that the former kind of action conflicts with strict or narrow (irremissible) duty, while the second kind conflicts only with broad (meritorious) duty. (*G,* p.424)

These two ways of failing the test are usually referred to as 'contradiction in conception' and 'contradiction in the will' respectively. It is unclear on the basis of the examples that he proceeds to give how Kant conceives of this contradiction test. However, this exegetical is dispute less urgent than

---

[35] See, *C2*, p.30: this is part of the reason why Kant ought to find it suspect to provide a separate derivation for the formula of humanity.

it might be, as the categorical imperative could not constitute an adequate canon for morally estimating actions on *any* of the readings of the contradiction test.

There are three broad strategies for reading the contradiction test; as a logical contradiction, as a teleological contradiction and as a practical contradiction.[36] The 'logical contradiction' interpretation is the one usually adopted by those less sympathetic to Kant (e.g. Hegel, Mill). It takes it that the procedure for applying the categorical imperative (hereafter CI procedure) is supposed to uncover a *logical* contradiction; on this reading, the reason why a maxim that contravenes a strict (perfect) duty cannot even be conceived as a universal law of nature is that to imagine such a maxim made universal is simultaneously to remove the conditions for the possibility of the original maxim. The best example to use here is that of the lying promise; such a maxim cannot be universalized because

> the universality of a law which says that anyone believing himself to be in difficulty could promise whatever he pleases with the intention of not keeping it would make promising itself and the end to be attained thereby quite impossible, inasmuch as no one would believe what was promised to him but would merely laugh at all such utterances as being vain pretences. (*G,* p.422)

Universalizing a permission to promise falsely would destroy the whole institution of promising; hence it would not be possible to simultaneously will both the maxim and its universalization, as the universalization destroys the institution of promising that the maxim itself depends on.

This interpretation suffers from having no account to give of the contradiction in the will: if the CI procedure is supposed to be defined in terms of logical contradiction, what do we make of Kant's claim, quoted above, that in cases of contradiction in the will 'this internal impossibility is indeed not found'?

---

[36] I borrow this schematization from Korsgaard (1985).

The teleological reading takes the contradictions in question to be provided by the natural purpose of human beings (use of their reason), or the natural purpose of human interrelations, which is the kingdom of ends. A contradiction might occur either because someone frustrated their own purpose *qua* human being, or because, if turned into a law of nature, it would frustrate the possibility of the kingdom of ends. Kant lays the ground for this argument with the argument about natural purposes in the *Groundwork* pp. 395-6, which starts from the premiss that

> In the natural constitution of a organized being, i.e. one suitably adapted to the purpose of life, let there be taken as a principle that in such a being no organ is to be found for any end unless it be the most fit and best adapted for that end. (*G,* p.395)

He then argues that it follows from this that our purpose is *not* our own happiness. His argument is as follows: happiness would be far more efficiently secured by a being that was guided merely by inclination *without* a will and practical reason. But given that we *do* have a will and are endowed with practical reason, it follows that their purpose cannot be the furtherance of our own happiness.

Indeed, his approach to the suicide example seems to suggest something like this; his argument against the person who feels that his life is no longer worth living ending his own life is as follows:

> One sees at once a contradiction in a system of nature whose law would destroy life by means of the very same feeling [sc. self-love] that acts so as to stimulate the furtherance of life, and hence there could be no existence as a system of nature. (*G,* p.422*)*

The point seems to be that i) the purpose of self-love and inclination in general is to ensure the furtherance of life; ii) hence if *from self-love* one commits suicide, one is acting against the natural purpose of self-love. The problem is that this reading of the categorical imperative, however plausible as exegesis of what Kant says in these particular passages of the

Groundwork, is in flat contradiction with the idea of autonomy. As we saw, any positing of an end for human beings that is *prior* to the moral law and that is to be made a determining ground of the will is explicitly ruled out by Kant's argument against perfectionism, quoted above. No such reading can therefore explain how pure reason can be practical.[37]

The third reading is that the contradiction in question is *practical;* that the agent would not be able to *act* on his maxim in a world in which the maxim were universalized. What is supposed to be tested here is whether you would be able to act on your maxim to achieve your purpose in a world in which it was a general law that people acted on that purpose. It is impossible to *act* on a maxim that presents a contradiction in conception (say, giving a lying promise) because in a world in which such behaviour was generalised it would be impossible to achieve what one intended through giving a lying promise, as promises would not be generally believed. As Korsgaard puts it, "What the test shows to be forbidden are just those actions whose efficacy in achieving their purposes depends on their being exceptional." (Korsgaard 1985, p.92)

This interpretation allows us to discriminate, in addition, between the contradiction in conception and the contradiction in the will. There is a contradiction in the will, on this reading, when a purpose to be pursued goes against an *essential* end of the will – so that, although a world in which acting on such a purpose were universal is possible, *willing* such a world contradicts some essential feature of the will.

---

[37] Kant expresses an understanding of the relationship between morality and teleology that is compatible with the autonomy later in the Groundwork:

'Teleology considers nature as a kingdom of ends; morals regards a possible kingdom of ends as a kingdom of nature. In the former the kingdom of ends is a theoretical idea for explaining what exists. In the latter it is a practical idea for bringing about what does not exist but can be made actual by our conduct, i.e. what can be actualized in accordance with this very idea.' (*G*, p.437n)

To paraphrase: morality requires that we act so as to make actual the ideals of morality, and so *make it the case that* the world *is* teleologically ordered. (cf: *G,* p.434: 'Morality consists in the relation of all action to that legislation whereby alone a kingdom of ends is possible.') Teleology, on the other hand, is a mode of explanation that *presupposes* that the world is (already) teleologically ordered. Morality does not proceed *from* teleology, but *towards* it. For an illuminating reconstruction of Kant's account of this, see Äpel (1997).

But there is a deep weakness that affects both this and the logical contradiction interpretation. There are many maxims that *ought* to be ruled out, but which are not ruled out by this test, regardless of whether we adopt the logical or the practical contradiction test.[38] The simple problem is this: there are many maxims that *ought* to be deemed impermissible, but which generate no contradiction, because there is nothing either logically or practically impossible about the maxim becoming a law of nature and the agent still acting on it in these circumstances.

Let us take as our example the maxim 'I will hunt down and kill anyone over the age of 30, in order to keep the population down'. Now, I take it that this *ought not* to be judged a morally permissible action by pure practical reason. But yet it is perfectly conceivable that a world should exist in which the results of following such a maxim held as a general law. (All we need imagine is that there are no human beings over the age of thirty and that people are eliminated as they reach their thirtieth birthday).[39] Hence this world passes the contradiction in conception test on the 'logical contradiction' interpretation.

It will also pass the contradiction in conception test on the 'practical contradiction' interpretation. For it will still be perfectly possible for the agent to act on the same maxim in this world. Indeed, it might be positively *required* of the agent that he act on this maxim in such a world.[40]

Can the contradiction in the will interpretation help us out here? It depends on how we interpret it. Korsgaard suggests the following:

> If a thwarted purpose is a practical contradiction, we must
> understand the contradiction in the will test this way: we

---

[38] I shall henceforth ignore the teleological contradiction interpretation, on the ground of its fundamental inconsistency with the idea of autonomy.

[39] The example here is (of course) indebted to the bad Seventies film, *Logan's Run.*

[40] I am assuming here that the agent willing the maxim is under the age of thirty. Otherwise things get a bit complex: would a 35 year old be able to act on this maxim in this world? (If so, would his first act have to be suicide?) However, any contradiction generated would be wholly inadequate to the purposes of morality; a restriction on murder that only applied to the over-thirties would be (to put it mildly) sub-optimal.

> must find some purpose or purposes which belong
> essentially to the will, and in the world where maxims that
> fail these tests are universal law, these essential purposes
> will be thwarted, because the means of achieving them will
> be unavailable. Examples of purposes that might be thought
> to be essential to the will are its general effectiveness in the
> pursuit of its ends, and its freedom to adopt and pursue new
> ends. (1985, p.96)

It is highly questionable whether there is any contradiction of the will's essential purposes involved in the 'Logan's Run' maxim if they are described in a way that is as formal as this. In fact, Kant seems to presuppose a 'thicker' account of the essential purposes of the will in the two examples of the contradiction in the will in the *Groundwork* – willing not to develop one's talents and willing never to help others. What seems to be contradicted is the requirement that we take humanity to be an *objective end.* (Note how these two examples correspond exactly to what he gives as the two ends of ethics in the *Metaphysics of Morals* – one's own perfection and the happiness of others.)

Though this makes sense of what Kant says and how he applies the test, it would lead to a circle in his account. Unless we take the end of humanity to be *separate from* the categorical imperative (which, we said, was incompatible with the structure of the *Critique of Practical Reason*), we cannot use the objective end of humanity to explicate the categorical imperative. For to do so would be to attempt to pull ourselves up by our own bootstraps. If we wish Kant's account to avoid this vicious circle, we can only require something completely formal of these essential features of the will; but such formal features are unable to exclude maxims such as the Logan's Run one.

This problem is serious. If it is followed out, it is rather disturbing for Kant's account of morality. Those actions such as violence, murder and rape, that morality should be strongest in its condemnation of, suffer no contradiction in conception if we attempt to will them as a universal law.

The attempts to convict them of a contradiction in the will remain dubious. In fact it seems that all the categorical imperative can do is to disbar those that depend on the existence of a particular institution and which if universalized would undermine that institution.[41] The claim that the categorical imperative provides a canon for morally estimating actions is therefore false.

Kant's second approach to the practicality of pure reason removes some problems by making the distinction between rules and judgment. Kant argues that this distinction is necessary in the *Critique of Pure Reason* and in the essay on *Theory and Practice*. He argues as follows that we need a faculty of judgment over and above any general rules we have.

i) No rule contains the conditions for its own application.

ii) If we supplement a rule with a second rule to tell us how to apply the first rule, then we will need a third rule to tell us how to apply the second rule, and so on ad infinitum.

iii) If there were this regress, then we would never be able to correctly apply a rule.

iv) On the assumption that we *do* correctly apply rules, then we must have a faculty of judgment 'whereby the practitioner distinguishes instances where the rule applies from those where it does not.'[42]

---

[41] I am indebted to Korsgaard's discussion of where the formula of universal law goes wrong. Korsgaard thinks that all is not lost for Kant however, as the formula of humanity can step in in such troublesome cases (1985, p.100). I have two points here: first (as I have already said) Kant cannot, consistent with his position in the *Critique of Practical Reason,* accept a separation between the formula of universal law and the formula of humanity. Second, once you admit that i) the formula of humanity and the formula of universal law proscribe different maxims, and ii) must be used in tandem so that they give a result that tolerably matches our ordinary moral beliefs, you have already given up on the claim that the categorical imperative is the *canon* for estimating actions and on the claim that pure reason can be practical. One would instead have transplanted the two formulations of categorical imperative into a general process of reflective equilibrium.

[42] 'On the Common Saying: "This May be True in Theory, but it does not Apply in Practice', p.61. See *C1,* A133/ B172: 'If [the understanding] sought to give instructions how we are to subsume under these rules, that is, to distinguish whether something does or does not come under them, that could only be by means of another rule. This in turn for the very reason that it is a rule, again

It follows that the task of moral philosophy is to give us rules, not to give us judgments in particular situations. But this 'purer' conception of the task of moral philosophy only serves to make manifest a deeper problem with the very idea of pure reason being practical.[43]

The problem is this: the formula of universal law, which as we have seen should be taken to be the fundamental form of the categorical imperative, depends on the ability of maxims to generate contradictions when they are universalized. But it follows from Kant's account of the categorical imperative that such contradictions cannot have any normative force unless the institution whose existence is responsible for the contradiction can be shown to be necessary a priori.

Even if we agree that a maxim could not be simultaneously willed with its universalization, then there is no reason to assign any normative force to this contradiction unless we take the institution to be *necessary*. Suppose, for example, that someone takes as his maxim 'I shall dress in such a way that I subvert the assumptions about manliness present in society'; now there is a contradiction in conceiving this as a universal law, because the maxim promises to *subvert* an institution (namely manliness), and if everyone aimed at such *subversion*, then there would no longer be any such conception of manliness to subvert. But unless this institution can *itself* be shown to be necessary by pure reason then the fact that the universalization of a maxim would undermine or destroy it has no normative force.[44]

Hence the relevance of Hegel's example of property. Now, if a society has an institution of property, then we can quite easily generate a contradiction

---

demands guidance from judgment. And thus it appears that, though understanding is capable of being instructed, and of being equipped with rules, judgment is a peculiar talent which can be practised only, and cannot be taught.'

[43] This problem as applying to the idea of pure reason's being practical as such will also affect any account that seeks to combine moral judgment with moral precepts.

[44] Indeed we could go further: if an institution is such that it *ought to be destroyed,* (slavery, for example) then acting on maxims that preserve the possibility of the institution may itself be culpable.

with the maxim 'I will give according to my ability, and take what I need'; for this would destroy the notion of property. But what needs to be shown is that the institution of property is necessary a priori. It will be only on this condition that any maxim that would dissolve it if universalized must be abandoned.[45]

This difficulty does not escape Kant's notice. In the *Metaphysics of Morals*, he draws a distinction between what he has done in the *Metaphysical Foundations of Natural Science* and what he believes to be necessary in morality:

> It has been shown elsewhere that for natural science, which has to do with objects of outer sense, one must have a priori principles, and that it is possible, indeed necessary, to prefix a system of these principles, called a metaphysical science of nature, to natural science applied to particular experiences, that is, to physics. Such principles must be derived from a priori grounds if they are to hold as universal in the strict sense. But physics (at least when it is a question of keeping its propositions free from error) can accept many principles as universal on the basis of evidence. ....
>
> But it is different with moral laws. They hold as laws only insofar as they can be *seen* to have an a priori basis and to be necessary. Indeed, concepts and judgments about ourselves and our deeds and omissions signify nothing moral if what they contain can be learned merely from experience. And should anyone let himself be led astray into making something from that source into a moral principle, he would run the risk of the grossest and most pernicious errors. (*MM,* pp. 214-5)

In other words, we will need to perform a transcendental deduction of each of these institutions. Whilst we can imagine this working for, perhaps,

---

[45] I take this argument to follow from the one that Kant uses in the *Critique of Practical Reason* to establish that moral obligation requires a categorical imperative that is a priori necessary (see above). Just as in that argument inclinations were debarred from serving as the basis for the moral law because they do not have a priori necessity, so here any institution that cannot be demonstrated to be necessary a priori must be debarred from serving as the basis for moral precepts. It follows that if anyone objects to this argument against Kant, they should really object to his initial argument that moral obligation requires a categorical imperative and moral laws that are given a priori.

promising and property,[46] the most basic institutions connected to morality, such as the family *cannot* be shown to be necessary a priori simply because they must be deemed empirical and contingent from the perspective of pure reason.

Pure reason thus fails to be practical. If used as a canon of moral judgment it will fail to pick out those maxims that are expressive of morality; even more seriously, the formula of universal law will not be able to generate moral precepts without relying on institutions that are themselves contingent from the perspective of pure practical reason. Pure reason as Kant conceives it is simply not capable of generating a way of deciding which maxims are morally permissible and what the moral precepts are that does not surrender the determination of the will to something outside pure reason.

It follows that Kant's account of morality makes moral obligation impossible: he argues, as we saw, in the *Critique of Practical Reason* that moral obligation is possible only if pure reason can be practical. But we have now shown that pure reason *cannot* be practical; hence Kant would have to admit that moral obligation is not possible. As we have seen, though, this impossibility does not affect morality as such: for morality as such need not make such inflated claims about the conditions under which moral obligation is possible.

Interestingly, it turns out that the very idea of pure practical reason is incoherent. Kant argues that normativity must go all the way down; he also argues that normativity is confined to the a priori. He is thus in trouble now

---

[46] Kant's argument for the necessity of property is, however, pretty feeble. He urges that it is a postulate of practical reason that 'It is possible for me to have an external object of my choice as mine, that is, a maxim by which, if it were to become a law, an object of choice would *in itself* (objectively) have to *belong to no one* (*res nullius*) is contrary to rights.' (*MM,* p.246) He attempts to show the necessity of this postulate a priori, because, if the contrary held, then there would be objects that could not be rightfully used by anyone, and 'freedom would be depriving itself of the use of its choice with regard to an object of choice, by putting *usable* objects beyond any possibility of being *used.'* (*MM,* p.246) This might have some force as an a priori argument against someone who claimed that certain things cannot be rightfully used by anyone. But this is not the point at issue: the alternative to property is rather the claim that *all* things are rightfully usable by everyone.

that it is clear that practical reason must be at least partly determined by grounds that are empirical; for it now appears that pure reason is *insufficient* to determine what is to be done, because it is insufficient to determine the relevant moral precepts. But for Kant, the idea of a reason that is not responsible for its own judgments, but rather acquiesces in having its mind made up for it is a nonsense:

> Now we cannot possibly think of a reason that consciously lets itself be directed from outside as regards its judgments; for in that case the subject would ascribe the determination of his faculty of judgment not to his reason, but to an impulse. (*G,* p.448)

It follows that (on good Kantian grounds) Kant's conception of pure practical reason is incoherent; for it transgresses the basic conception of what it is for a being to *have* reason.

2.4 *Why Kant's account of morality gets human life wrong*

What I say about Kant's conception of human life will be overwhelmingly negative; though it goes without saying that I do not think that is unfairly so. I should perhaps explain in advance why I shall not address in any detail the neo-Kantian attempts to get Kant off the hook on various of the charges I shall make.

My reasons for doing this fall into two camps: first, and perhaps paradoxically, they come from a respect for Kant's greatness as a systematic philosopher. Kant's way of understanding human ethical life follows *directly* from his understanding of the morality, and his account of morality forms a highly integrated system, in which the various parts are supposed to necessitate one another. Once we understand this, we see that there is much less room for manoeuvre in interpreting him than we might like. We cannot, for instance accept some parts of his theory without accepting all of it, unless we have explicit arguments *against* Kant to show that the entailments between different parts of his theory do not in fact

hold. As we have already seen, it is the a prioristic conception of the moral law which forms the centrepiece of this theory; and it is this conception of the moral law which causes the problems.

Second, our interest in this thesis is in the tradition of morality, not in Kant *per se:* reading Kant in this strict way is very helpful for our broader project, as it both allows us to see the ways in which the tradition of morality can and has failed to live up to its own best insights, and provides a valuable context for the Nietzschean assault on morality in the next chapter.

Kant's conception of human life follows from his account of the moral law. I shall argue that this theory of human life fails both on a factual level, and presents a rather stunted expression of the sort of life that morality should involve. I shall begin by explaining what this theory is, and then explain why it gets human life wrong.

This account of the moral law is, as we have seen, first of all a prioristic. The following two passages are indicative of his general position:

> Ethical duties must not be determined in accordance with the capacity to fulfil the law that is ascribed to man; on the contrary, man's moral capacity must be estimated by the law, which commands categorically, and so in accordance with our rational knowledge of what men ought to be in keeping with the Idea of humanity, not in accordance with the empirical knowledge we have of men as they are. (*MM,* p.405)

> The teachings of morality... command for everyone, without taking account of his inclinations, merely because and insofar as he is free and has practical reason. He does not derive instruction in its laws from observing himself and his animal nature or from perceiving the ways of the world, what happens and how men behave... Instead, reason commands how men are to act even though no example of this could be found.[47] (*MM,* p.216)

---

[47] See also, for example, *G,* p.425: 'Duty has to be a practical, unconditioned necessity of action; hence it must hold for all rational beings (to whom alone an imperative is at all applicable) and for this reason only can it also be a law for all human wills.'

For Kant, the moral law plays a *constitutive* role in human life: it sets up the conditions under which human life is to be lived, and qua constitutive it is itself beyond criticism. If we bear this in mind, then much of the rest of Kant's conception of human life follows automatically.

The most important result that follows is that there is an automatic 'fit' between the moral law and the world: while we are not to look to man to define what our duty is, it is nonetheless true that it is always *possible* for each person to do their duty at each time. Kant explains how this is possible in terms of his i) theory of transcendental freedom, which guarantees that, whatever habits a person may have formed, and whatever their upbringing, they are still absolutely free to do what is morally required, and ii) an a priori thesis which legislates that genuine moral dilemmas are impossible, since

> duty and obligation are concepts that express the objective practical *necessity* of certain actions and two rules opposed to each other cannot be necessary at the same time, if it is a duty to act in accordance with one rule, to act in accordance with the opposite rule is not a duty but even contrary to duty; so a *collision of duties* and obligations is inconceivable (*obligationes non colliduntur*). (*MM,* p.224)

Hence we can know a priori that it will never happen that by an unfortunate accident a person will finds themselves forced to do something that is morally wrong.[48]

Now, if it is always possible to do what is right and moral dilemmas are impossible it follows that it is always fair (and, indeed, morally required) to blame someone if they fail to do their duty.

---

[48] Kant's reasoning continues: 'However, a subject may have, in a rule he prescribes to himself two *grounds* of obligation (*rationes obligandi*), one or the other of which is not sufficient to put him under obligation (*rationes obligandi non obligandes*), so that one of them is not a duty. When two such grounds conflict with each other, practical philosophy says, not that the stronger obligation takes precedence (*fortior obligatio vincit*), but that the stronger *ground of obligation* prevails (*fortior obligandi ratio vincit*).' (*MM,* p.224)

Actions and character too are to be measured from the perspective of the moral law. On Kant's account, what matters, morally speaking, is not the result of the action – it is not what is brought about by it – but the maxim *from which* the agent acts:

> An action done from duty has its moral worth, not in the purpose that is to be attained by it, but in the maxim according to which the action is determined. The moral worth depends, therefore, not on the realization of the object of the action, but merely on the principle of volition according to which, without regard to any objects of the faculty of desire, the action has been done. (*G,* pp. 399-400)

The only motive which it is morally worthy to act from is respect for the moral law: it follows that all other motives are deprecated. Only actions done *from* duty (that is, respect for the moral law) have moral worth; otherwise, they may have *legality* (that is, outward conformity with the moral law), but not morality. Acting from emotions such as love or compassion has no moral worth.[49]

Kant defines the good life for a human being *in terms of* the moral law. Moral virtue is the *supreme* good for a human being: all morally virtuous lives enjoy a *lexical priority* over all non-virtuous ones. But moral virtue is not the *complete* good, for it possible to imagine a life that is *better* than one that is merely morally virtuous, namely one that in which that person was also happy in proportion to their virtue. Such a life would be the *complete* good for a human being.

The moral law also pervades his account of evil. According to Kant, human beings stray from morality because their sensuous nature brings with it inclinations, which aim at natural good (that is, our own happiness), not at what is required by morality.[50] Although there is nothing wrong with these

---

[49] 'It is of the utmost importance in all moral judging to pay strictest attention to the subjective principle of every maxim, so that all the morality of actions may be placed in their necessity from duty and from respect for the law, and not from love or leaning toward that which the action is to produce.' (*C2,* p.81)

[50] I have found Silber (1960) helpful on this point.

inclinations *per se,* they tempt us to subordinate the claims of morality to them; evil just is the subordination of the moral law to inclination in choice.[51] Immorality is always a matter of making oneself an exception to a general rule *for the sake* of one's inclinations, which in a sense is understandable (for we cannot help being moved by our inclinations), but it is definitely not excusable.[52]

Taken together, we get the following picture: we are to live in accordance with immutable standards; to live in such a way is always possible; and to live dutifully in accordance with such standards is the supreme human good.

2.4.1 *Some criticisms of Kant's worldview*

By presupposing this a prioristic conception of moral rules, Kant gives himself carte blanche simply to ignore any complaints that might arise from a deeper or broader sensitivity to the contours of human life. For it will follow a priori that *no* empirical understanding of human beings and of human life could in any way alter or have anything to say about morality.

However, we should not let Kant get away with this, for two reasons: first, it is a question-begging manoeuvre. Kant attempts to forestall criticisms of his theory of human life on the grounds of empirical knowledge of human nature by simply ruling any such challenges out of court in principle. But given that the critic is arguing *from* empirical claims *to* the inadequacy of

---

[51] See for example, *Religion within the limits of reason alone,* where he says 'Natural inclinations, *considered in themselves* are *good,* that is, not a matter of reproach, and it is not only futile to want to extirpate them but to do so would also be harmful and blameworthy.' (p.51) For Kant, 'only what is opposed to the moral law is evil in itself' (ibid); hence human evil consists in *subordinating* the moral law to the goods aimed at by sensuous nature.

[52] See also *G,* p.424: 'If we now attend to ourselves in any transgression of a duty, we find that we actually do not will that our maxim should become a universal law – because this is impossible for us – but rather that the opposite of this maxim should remain a law universally. We only take the liberty of making an exception to the law for ourselves (or just for this one time) to the advantage of our inclination.' (This might cited as further evidence that the coherence of Kant's account of morality requires the formula of humanity and formula of universal law to coincide.)

Kant's theory of human life, it is no reply to the critic to argue that *given that* the moral law is given a priori, the critic's charge has no weight.

Second, we saw in 2.3 that moral obligation is in fact impossible if we follow Kant's presuppositions about the moral law, and in 2.2 that there does not seem to be any very good reason for following him in his assumption that these presuppositions are necessary for morality, and so we should take the very idea of an a prioristic account of morality to be in the dock, and certainly should not presuppose it.

Hence in what follows I shall take it that we should test the a priori claims that Kant makes against our empirical reality, and that if what he says seems to offend against the unavoidable structures of human life, that this is a problem for his account, rather than a problem with human beings.

Just about every aspect of Kant's world view as we have just explained it is deeply problematic.

The most fundamental problem is Kant's idea of the moral law itself: as we have seen this is foundational to Kant's account. Put simply, it is simply not a live option any more to believe in the moral law in the way Kant does. This is not to deny that Kant himself took the belief that seriously, but simply to register the fact that we cannot. We can see this most obviously when we see how badly Kant's attempts backfire when he raises to the heights of his rhetoric to expound the majesty of the moral law:

> Duty! Thou sublime and mighty name that dost embrace nothing charming or insinuating but requirest submission and yet seekest not to move the will by threatening aught that would arouse natural aversion or terror, but only holdest forth a law which of itself finds entrance into the mind and yet gains reluctant reverence (though not always obedience) – a law before which all inclinations are mute even though they secretly work against it: what origin is worthy of thy noble descent which proudly rejects all kinship with the inclinations and from which to be

descended is the indispensable condition of the only worth
which men alone can give themselves? (*C2,* p.86)

This strikes me as comical in its bathos; and I have never yet met anyone
who is rendered suitably awestruck by it.

It is not open to Kant to argue that this shift in consciousness is a symptom
of our moral decline, because if his theory were true, such moral decline
would, in fact be impossible. If his account were true, we would no more
be able to escape thinking of the moral law in this way than we would be
able to escape thinking of ourselves as existing in time or thinking of the
world as ordered according to causal laws.

Kant, we may assume, correctly describes the phenomenology of his own
moral experience, but goes badly wrong in assuming that human ethical
experience *must* have this form: what we have instead is no more than 'an
audacious generalization on the basis of very narrow, very personal, very
human, all too human facts' (Nietzsche, 1886, Preface)

In fact there have always been many people for whom Kant's theory of the
necessary structure of the human will seems just plain false, both those
who are morally good, and those who are morally bad.

From the perspective of the morally good, there are many who are critical
of any position that places rules before human beings, and which effaces
the differences between human beings. For them, it is human beings that
matter, not rules: the rules must exist for our benefit, not vice versa.[53] We
must look again and see what the differences are between people and
between peoples; we must notice that persons are different in ways that
render problematic the very idea of moral precepts.[54]

---

[53] A thought that first gains entrance to the tradition of morality through the ethics of Jesus, I think
(see Mark 2.27: 'The Sabbath was made for man, not man for the Sabbath.')

[54] See for example, George Eliot: "All people of broad, strong sense have an instinctive
repugnance to the men of maxims; because such people early discern that the mysterious complexity
of our life is not to be embraced by maxims, and that to lace ourselves up in formulas of that sort is
to repress all the divine promptings and inspirations that spring from growing insight and sympathy.
And the man of maxims is the popular representative of the minds that are guided in their judgement
solely by general rules, thinking that these will lead them to justice by a ready-made patent method,

The case of the morally bad is particularly interesting: Kant's account of the moral law forces him to assume that even the worst person must constantly have the moral law before them. The only space that he can find for evil in his account, as we saw, is to declare that evil consists in placing self-love (one's own happiness) above the moral law.[55] Whilst this obviously fits some cases, it is desperately implausible as a theory of human wrong-doing, because it rules out a priori two ways of living a human life that, though destructive are nonetheless common: first, the ethical code held by most career criminals, where one's pursuit of one's own happiness or self-interest is in no way bound by moral considerations; and second, sheer wanton malevolence, bounded by neither self-love or the moral law, a possibility which unfortunately appears to be all too real in human life.[56]

Once we have broken the spell of Kant's account of the moral law, we can see that there is something deeply problematic in the attempt to construct a moral theory that applies to rational beings as such and which applies only as a side effect to human beings. For such an account can only take into account those features of human beings that are entailed in what Kant considers to be pure practical reason. This has the unfortunate effect of making the account peculiarly unreceptive to the relationships that human beings the world over value most, such as the bond between parents and child, true friends, and relationships between lovers: all of these come to

---

without the trouble of exerting patience, discrimination impartiality – without any care to assure themselves whether they have the insight that comes from a hardly-earned estimate of temptation, or from a life vivid and intense enough to have created a wide fellow-feeling with all that is human.' (1860, Book VII; The Final Rescue II, para. 29)

[55] In the *Religion,* Kant does briefly canvass the possibility of a 'reason exempt from the moral law, a *malignant reason* as it were', but then states that such a being would be 'devilish', and that man is not such a creature (p.30); he addresses the issue again at *MM* p.320n, where he states that 'As far as we can see, it is impossible for man to commit a crime of this kind, a formally evil (wholly pointless) crime'. The reasons why Kant cannot allow the possibility of human 'devilishness' is that this would disrupt the morality vs self-love dichotomy that (as we saw) plays a key role in the *Critique of Practical Reason* account, and second, that it would force him to admit that the moral law is more contingent to the human makeup than he would like it to be.

[56] Richer and more psychologically plausible accounts of human badness are provided by Glover 1999 and Oppenheimer 1996.

seem to be dangerous and to require a watchfulness lest one overstep the proper boundaries and come closer than a proper respect demands.[57]

This problem is exacerbated by making duty the sole morally worthy motive:[58] if we emphasize duty then we will misunderstand the most basic human ethical interactions – the parent-child bond; friendship; sexual love, in fact all those elements which are the cradle of moral goodness. We prefer to be treated from motives other than duty by those people we care about (few people would be happy to discover that their parents had acted *merely* from duty in bringing them up); and in turn become worried ourselves if we find nothing but duty to motivate us in our encounters with them.

Further, this conception of moral motivation sits very ill with our common sense judgment about those cases in which we are most certain of attaching moral worth: cases of supererogatory or heroic action, where a person goes far beyond what could be expected of a rational being in general (and hence beyond their duty). Further, whilst such actions *may* conceivably be done from respect for the moral law, rather more often they are done spontaneously or 'because it was the only thing to do in the situation'.

But perhaps most important, there are systematic problems that attend to the living life for the sake of duty alone: what happens to the agent's own life when he is reduced to acting all the time from duty, and disregarding anything else that he might deem worthwhile to the extent that it interferes with this goal?

---

[57] Kant theorizes friendship as a struggle between two forces – love which brings the friends closer together and respect, which forces them to keep a proper distance from one another, and he opines that 'even the best of friends should not makes themselves too familiar with each other' (*MM*, p. 470), on the grounds that 'Although it is sweet to feel in such possession of each other as approaches fusion into one person, friendship is something so delicate (*teneritas amicitiae*) that it is never safe for a moment from interruptions if it is allowed to rest of feelings, and if this mutual sympathy and self-surrender are not subjected to principles or rules preventing excessive familiarity and limiting mutual love by requirements of respect.' (*MM*, p. 471)

[58] I do not dispute that duty must play an indispensable role in morality, as I said in chapter one; but to make duty the sole morally worthy motive is a deep mistake.

Kant sees this problem and thinks that it is serious. He frames it as follows: first, as we saw, he conceives of moral virtue as the supreme good, though not the complete good; he conceives of virtue as worthiness to be happy and hence lack of moral virtue as worthiness to be unhappy.[59] The complete good would then be virtue and happiness proportionate to this virtue. The problem then is that although happiness *ought* to be proportional to virtue, in the world as we experience it, this is far from being the case. But given that morality presupposes this, there must then be an afterlife in which this deficiency is made up and a God who sees to it that the good finally get their reward and the bad are finally punished.

Bringing God in by the back door in this way as a 'postulate of practical reason' is highly implausible: but the fact that Kant thinks we *need* God to explain how the life of moral goodness can make sense tells us that there are important mistakes in his system.

The mistakes are as follows: first, it is a mistake to think of moral virtue as the supreme good: our own welfare cannot be that self-contained. It is obvious that things may happen to me or to those I care about which render laughable the claim that, in virtue of my moral virtue, I have in my hands an enviable life. We human beings are not invulnerable.[60]

Second, it is a mistake to attempt to hold moral virtue apart from happiness: we do better if we bring the pursuit of morality closer to the pursuit of a humanly fulfilling life by making the morally virtuous life one that incorporates within it those activities and relationships of mutual recognition which make for valuable human life.[61]

---

[59] Kant thus believes firmly in retributive punishment: "When, however, someone who delights in annoying and vexing peace-loving folk receives at last a right good beating, the beating is certainly a bad thing, but everyone approves of it and considers it as good in itself even if nothing further results from it; nay, even he who gets the beating must acknowledge, in his reason that justice has been done to him, because he sees the connection between well-being and well-doing, which reason inevitably holds before him, here put into practice." (*C2*, p.61)

[60] I discuss this claim further in chapter seven.

[61] I discuss these claims further in chapter six.

Third, it is a mistake to insist that that virtue should be proportional to happiness. We can simply give up on the idea of virtue as worthiness to be happy: we can still argue that there are certain goods that are only available to those that are morally good and are realised in morally decent relationships, without thinking that there has to be an *additional* reward for the good and an *additional* punishment for the bad. If we have confidence in the life of moral goodness, we can happily admit that the good person's 'reward' is nothing other than being the sort of person he is, and the bad person's 'punishment' no more than being the sort of person he is.

2.5 *Conclusion*

After the breakdown of the credibility of Kant's conception of the moral law, we must start again with morality from the question of why we should take morality seriously. We take this up in the next chapter by looking at Nietzsche's critique of morality. As we shall see, through his account of the dignity of human beings, Kant has the beginnings of an answer to this problem, even though it unfortunately remains largely hidden in his theory of human life.

But what then of the great positives of Kant's account of morality with which I started this chapter? These will be retained as far as possible and redeployed in the gradually changing framework we develop throughout this work.

Here it might be appropriate to remember something that Kant says about his own system and its relation to previous philosophy; as he suggests previous philosophy should be treated, so must we treat him:

> Yet since, considered objectively, there can only be one human reason, there cannot be many philosophies; in other words, there can only be one true system of philosophy from principles, in however many different and even conflicting ways men have philosophized about one and the same proposition. ... Although the *new system* excludes all

others, it does not detract from the merits of earlier moralists ... since without their discoveries and even their unsuccessful attempts we should not have attained that unity of the true principle which unifies the whole of philosophy into one system. (*MM,* p.207)

We can admire Kant's ambition and his subtlety in attempting to carry out his purpose; but we should not let respect for the man stop us from doing what is philosophically necessary – putting his whole system of moral philosophy to the test of reason. Now that this system has been found wanting, we need to find a new framework within which to construct a moral philosophy that is both internally consistent and fits the essential structure of human life. In this Kant will be of great value; but only after we start to build a new house to a new design will we find that we are able to reuse the raw materials that the destruction of Kant's account of morality has left us.[62]

---

[62] Perhaps a less charitable critic than myself might remind me of something Pa Ubu says in *Ubu Enchained*: "Hornstrumpot! We shall not have succeeded in demolishing everything unless we demolish the ruins as well. But the only way I can see of doing that is to use them to put up a lot of fine, well designed buildings."

PART TWO:

DIGNITY

MAP OF PART TWO

Chapter three, *Nietzsche's Critique and the Necessity of Affirmation*, begins with Nietzsche's critique of morality. Nietzsche objects to morality on three basic grounds: it attempts to legislate one ethical system as appropriate for everyone; it attempts to make everyone equal rather than respecting the distinctions of *rank* between human beings; and it is based on a resentment of other more vital modes of living rather than on the positive value of moral goodness.

I argue that the idea of the dignity of human beings is morality's best resource for answering these charges: through it we can explain why one mode of life is appropriate for all; why all human beings are equal and why morality in fact has a positive and affirmative value at its base.

The rest of part two examines the possibility of justifying the claim that all human beings have a dignity.

Chapter four, *Korsgaard's Argument for the Normativity of Morality*, examines Christine Korsgaard's complex and subtle argument for the normativity of morality. This aims to show that the form of reflective self-consciousness we have leads, by an inexorable chain of reasoning, to the necessity of us valuing ourselves as moral beings with a dignity and valuing all other human beings in the same way. I reconstruct the argument in some detail, but finds that it breaks down at the crucial point: the move to the claim that we *must* value our humanity and thus attribute a dignity to ourselves.

Chapter five, *Charles Taylor's Ethics of the Good,* looks at and extrapolates further from Charles Taylor's writings. The argument I reconstruct aims to show that i) we have a basic need to find our lives meaningful; ii) this need

must be expressed in finding some things *good*; iii) where this sense of the good does not include the affirmation of the dignity of human beings, it is a result of a form of self-deception or unwillingness to see on the agent's part; iv) therefore if an agent sees things rightly, she will perceive the dignity of human beings and act accordingly. I argue that whilst this account is promising, we simply have no way of demonstrating either iii) or iv).

Chapter Three

Nietzsche's Critique and the Necessity of Affirmation

3.1 *Introduction*

In moving over to part two of this thesis, we cross over a dividing line between the pre-Nietzschean and the post-Nietzschean approaches to morality. We saw in the first chapter how Nietzsche's approach to morality attempts to do two things simultaneously: first to demonstrate the incoherence of the account morality gives of itself, and second to interpret morality in way that makes it coherent once more, but which places it in a bad light.

Any post-Nietzschean account of morality needs to be able to withstand the squeeze from both of these sides: it must be both coherent, and find a way of deflecting the criticisms Nietzsche makes. The positions we shall be looking at in Parts Two and Three are designed to do precisely this.

3.2 *Nietzsche's interpretation of morality*

Nietzsche's interpretation of morality[63] presupposes, following ancient ethics, that the fundamental standard by which ethical systems should be judged is the flourishing of the individual. Where it differs from these ancient accounts is that it does not presuppose that there are *universal* standards for human flourishing, and on the contrary thinks that it is a dangerous mistake to apply the same standards to everyone:

---

[63] Here and elsewhere when 'morality' is used unaccompanied, I intend it to be taken in the way I explicated it in chapter one, not in Nietzsche's usage of the term. Later, when I refer to 'master morality' and 'slave morality' these will be following Nietzsche's usage.

> Every unegoistic morality which takes itself as unconditional and addresses itself to everybody is not merely a sin against taste: it is an instigation to sins of omission, one seduction *more* under the mask of philanthropy – and a seduction precisely for the higher, rarer, privileged. Moralities must first of all be forced to bow before *order of rank,* their presumption must be brought home to them – until they at last come to understand that it is *immoral* to say: 'What is good for one is good for another.' (1886, §221)

What is worse, the values morality attempts to make universal are those that are suitable only for weak people, and so universalizing them is harmful above all to the strong. This is an important problem because the sort of values that make human life worth living can only be created by the strong, and the strong can only create such values in situations where distinctions of rank are observed (in Nietzsche's terms, out of a *pathos of distance*). All in all the charge is that in moving away from *elitist* values, where one values according to one's rank as a human being, towards *universalistic* and *democratic* values, morality sacrifices the chance of greatness and diminishes mankind, reducing everything to the level of a dull, insipid mediocrity.[64]

For Nietzsche, morality is an *inversion* of the proper order of values: as such there is something strange and problematic about it. How did we come to believe it in the first place? Nietzsche's suggestion is that we see the sort of mode of valuation involved in morality as powerlessness becoming creative and expressing itself in resentment at the uninhibited powerful outflowing of vitality that belongs to the elite. By sheer cunning this mode of valuation has insinuated itself into general consciousness to such an extent that even the strong feel guilty in exercising their strength.[65]

---

[64] "What if a regressive trait lurked in 'the good man', likewise a danger, an enticement, poison, a narcotic, so that the present *lived at the expense of the future*? Perhaps in more comfort and less danger, but also in a smaller-minded, meaner manner? . . . So that morality itself were to blame if man, as species, never reached his *highest potential power and splendour*? So that morality itself was the danger of dangers? …." (1887, Preface, 6)

[65] Nietzsche argues (1887, I.13) that the invention of freedom of the will has been the major way in which this inversion has been managed: "no wonder then, if the entrenched, secretly smouldering

Nietzsche expresses this theory in the first essay of the *Genealogy of Morals*. His foundational claim is that valuation starts with those who are pre-eminent seeing themselves as vastly superior to the mass of common riff-raff and praising themselves as superior.[66] It is precisely out of the awareness of the vast gulf between themselves and the others (in Nietzsche's terms, this *pathos of distance*) that 'they claimed the right to create values and give these values names.' (1887, I.2)

On this mode of valuation 'good' enjoys a conceptual priority to 'bad'. The good first of all pick themselves out *as* the good, and the bad are merely those who are left over. 'Bad' is only ever a shadowy existence, an after thought.[67] But what about the naturally weak? They cannot affirm themselves as the weak in the face of those who are naturally stronger. Rather, they 'project feelings of hatred and vengefulness onto the noble caste, feelings which, however, they cannot actually act upon, precisely because of their inferiority, and therefore take refuge in imaginary revenge. These revenge fantasies harden into *ressentiment* – into jealousy of the superior individuals' qualities and suspicion of everything that makes their superiority manifest.' (Joas, 2000, p.24)

Nietzsche calls the first mode of valuation 'master morality' and the second, 'slave morality'. Where master morality is affirmative – the masters affirm themselves as 'the good', slave morality is by its very nature *reactive:*

> Slave morality says 'no' on principle to everything that is
> 'outside', 'other', non-self': and *this* 'no' is its creative

---

emotions of revenge and hatred put this belief to their own use and, in fact, do not defend any belief more passionately than that *the strong are free* to be weak, and the bird of prey are free to be lambs: - in this way, they gain the right to make the birds of prey *responsible* for being birds of prey…" See also Williams (1993) on this.

[66] "The judgment 'good' does *not* emanate from those to whom goodness is shown! Instead it has been 'the good' themselves, meaning the noble, the mighty, the high-placed and the high-minded, who saw and judged themselves and their actions as good, I mean first-rate, in contrast to everything lowly, low-minded, common and plebeian." (Nietzsche 1887, I.2)

[67] '[I]ts negative concept 'low', 'common', 'bad' is only a pale contrast created after the event compared to its positive basic concept, saturated with life and passion. "We the noble, the good, the beautiful and the happy!"' (Nietzsche, 1887, I.10)

deed. This reversal of the evaluating glance – this *inevitable* orientation to the outside instead of back onto itself – is a feature of *ressentiment*: in order to come about, slave morality first has to have an opposing, external world, it needs, physiologically speaking, external stimuli in order to act at all, – its action is basically a reaction. (1887, I.10)

Whereas in a master morality 'good' has a conceptual priority and 'bad' is an after-thought, a slave morality must make what it is opposed to – which it denominates 'evil' – conceptually prior, and make its 'good' an after-thought. And who is 'evil' on the slave morality's account? 'The stern reply is: *precisely* the 'good' person of the other morality, the noble, powerful, dominating man, but re-touched, re-interpreted and re-viewed through the poisonous eyes of *ressentiment.*' (1887, I.11)

Putting all these thoughts together, Nietzsche expresses his fundamental charge against morality in the following way:

> To refrain from mutual injury, mutual violence, mutual exploitation, to equate one's own will with that of another: this may in a certain rough sense become good manners between individuals if the conditions for it are present (namely if their strength and value standards are in fact similar and they both belong to *one* body). As soon as there is a desire to take this principle further, however, and if possible even as the *fundamental principle of society,* it at once reveals itself for what it is: as the will to the *denial* of life, as the principle of dissolution and decay. Even that body within which, as was previously assumed, individuals treat one another as equals – this happens in every healthy aristocracy – must, if it is a living and not a decaying body, itself do all that to other bodies which the individuals within it refrain from doing to one another: it will have to be the will to power incarnate, it will want to grow, expand, draw to itself, gain ascendancy – not out of any morality or immorality, but because it *lives,* and because life *is* will to power. … 'Exploitation' does not pertain to a corrupt or imperfect or primitive society: it pertains to the *essence* of the living thing as a fundamental organic function, it is a consequence of the intrinsic will to power which is precisely the will of life. – Granted this is a novelty as a

theory – as a reality it is the *primordial fact* of all history:
let us be at least that honest with ourselves!' (1886, §259)

To sum up, Nietzsche's charge is that the tradition of morality, is flawed, first in its very idea of setting universal ethical standards for human beings; second in its insistence on equality and its attempt to efface distinctions of rank; and third because it is a slave morality, powered by a resentment of others, whom it denominates evil, and not by the life affirming power of its conception of goodness. Nietzsche is interested, fundamentally, in what makes life worth living and his charge is that the general regulation of society according to the standards of morality stands in the way of this.

It is best to begin by repaying Nietzsche's perennial honesty in like kind, and admit that Nietzsche's attack is not totally without foundation; and to simply accept with gratitude his uncovering of the myriad occasions in which morality has through its history indeed had a corrosive effect on the lives both of those who live according to it, and those who interact with them.

But it no way follows that morality must *of necessity* lead to these negative effects. Rather it is open to us to take Nietzsche's analyses as a *challenge*. We can agree that *if* morality can only be motivated by sickly resentment, then it is indeed indefensible: but the door is open to us to explain how morality might make life worth living for each and every person.[68]

Morality's most obvious resource for doing this is to invoke the dignity of human beings. The dignity of human beings, as it has been understood in the tradition of morality, undercuts Nietzsche's complaints about universal standards and about equality: as we saw in chapter one, morality has taken

---

[68] As Charles Taylor puts it: 'Nietzsche's challenge is based on a deep insight. If morality can only be powered negatively, where there can be no such thing as beneficence powered by an affirmation of the recipient as a being of value, then pity is destructive to the giver and degrading to the receiver, and the ethic of benevolence may indeed be indefensible. Nietzsche's challenge is on the deepest level, because he is looking precisely for what can release such an affirmation of being. His unsettling conclusion is that it is the ethic of benevolence which stands in the way of it. Only if there is such a thing as agape, or one of the secular claimants to its succession, is Nietzsche wrong.' (1989, p.516)

all human beings to have a dignity, and this dignity to make them equal in worth to one another, and to require a respect for each that can only properly be captured in universal rules.

Indeed, the dignity of human beings, one could argue, has always been the positive affirmation of worth at the heart of morality: it is worthwhile being moral simply because human beings – oneself and others – have a dignity that merits respect. The idea of the dignity of human beings, in other words, instantiates the sort of affirmation of life that Nietzsche looks to the strong to create from their pathos of distance: but instead of coming from imposing one's will and one's interpretations on the world, this affirmation comes from recognizing the value of human beings, oneself included.

Any account of morality that takes the dignity of human beings seriously is very far indeed from being a *will to nothingness,* a 'practical nihilism'.[69] (Nietzsche 1895, §7) Such an account of morality will, in fact, be a great bulwark *against* nihilism. The assertion of the dignity of human beings thus constitutes an act of affirmation of the value and worthwhileness of human life on a huge scale: an act of communal affirmation, through which each person affirms the value both of themselves and all other human beings.[70]

---

[69] This charge is levelled (not unfairly) at Schopenhauer's (1841) account of morality, though this should invite the question *why* he takes Schopenhauer's position to be the logical conclusion of 'Christian morality'.

[70] If we place the dignity of human beings at the forefront of our account of morality in this way, we get a different conception of moral rules from Kant's: we retain the idea of universal moral rules, but take them to follow as a corollary of the affirmation of human beings involved in according them a dignity. Paul Ricoeur explains the structure of such an ethics better than anyone: '[Y]ou shall not take life, you shall not steal, you shall not kill, you shall not torture. In each case, morality replies to violence. And if the commandment cannot do otherwise than to take the form of a prohibition, this is precisely because of evil: to all the figures of evil responds the *no* of morality. On the level of the ethical aim, however, solicitude, as the mutual exchange of self-esteems, is affirmative through and through. This affirmation, which can well be termed original, is the hidden soul of the prohibition. It is what, ultimately, arms our indignation, that is, our rejection of *indignities* inflicted on others.' (1992, p.221)

3.3 *The very idea of dignity*

Reading Nietzsche presents us with two sorts of problem for this approach to morality. The first problem is that on Nietzsche's view, the dignity of human beings is a human invention, and moreover one that works from factually false premisses:

> *The four errors.* – Man has been educated by his errors. First, he always saw himself only incompletely; second, he endowed himself with fictitious attributes; third, he placed himself in a false order of rank in relation to animals and nature; fourth, he invented ever new tables of goods and always accepted them for a time as eternal and unconditional: as a result of this, now one and now another human impulse and state held first place and was ennobled because it was esteemed so highly. If we removed the effects of these four errors, we should also remove humanity, humaneness, and "human dignity." (1882, §115)

Now, from Nietzsche's perspective, the fact that something is false is not in itself an objection to it: but those who have believed in the dignity of human beings have generally been rather more absolutist in their approach to truth. For them it has been important to say that it is *true* that all human beings have a dignity, and to provide arguments in support of this claim. So, even though this does not represent a serious objection to the idea of the dignity of human beings from *Nietzsche's* perspective, it is one that morality will want to take seriously.

The second problem is, as we have seen, Nietzsche's contention that distinctions of rank are a prerequisite for vital human life and hence that it is a mistake to attempt to install a table of values in which everyone is equal.

I shall address the first in the remainder of this chapter. The second must wait until chapter six for a proper consideration.

We could reply to the first problem Nietzsche raises in two ways.[71] The first reply would argue that *pace* Nietzsche, the affirmation of human beings required by the concept of dignity *does* latch onto something real: there is an identifiable special trait or traits that all human beings have that makes it appropriate to accord them a dignity.[72]

The second reply would be more circumspect, and possibly more conciliatory to Nietzsche's way of proceeding. It would eschew the claim that there is an identifiable special trait or traits that human beings have *in virtue of which* it is appropriate to accord them a dignity. It will suggest that we only need defend the *attitude of affirmation of other human beings,* and that the defence of such an attitude need not require there to be special value-inducing traits of human beings that are identifiable separately from our attitude of affirmation.[73]

I shall spend the rest of this chapter examining the possibility of giving a positive justification, before concluding that the possibility of so doing seems remote, and that the sceptical justification seems to be much the best way to proceed. The next two chapters then examine Christine Korsgaard's and Charles Taylor's unsuccessful attempts to make something out of the sceptical justification: I give my account of the dignity of human beings in chapter seven and eight.

3.4 *The positive justification*

When we introduced the topic of the dignity of human beings in chapter one, we said that it could be decomposed into the following three thoughts:

---

[71] Margalit, whose work I have found helpful has a similar discussion, on the topic of how to justify respect for human beings. He also considers a third option, which he calls the 'negative justification', which would not attempt to 'provide a justification for respecting people, but only for not humiliating them.' (1996, p.84) I leave this on one side as it does not even aspire to give the sort of affirmation which we require of human life.

[72] Margalit calls this a 'positive justification'

[73] Margalit calls this a 'sceptical justification': In his words, it 'giv[es] up the search for a justifying trait that is prior to the attitude of respect. Instead, the attitude of respect becomes the starting point, while the respect-evoking trait of being human is derived from this attitude itself.' (1996, p.77)

1. Human beings as such have a worth that is categorially higher than anything non-human (animals, things etc.).

2. Human beings are equal in worth to one another.

3. The dignity ('absolute inner worth') of human beings entails that we treat them with *respect*.

There are three questions to be asked: first, what sort of a value is a dignity; second, how *could* a special trait bestow dignity on its possessor; third, are there any special traits that could at all plausibly fit this bill?

I shall consider each in turn.

3.4.1 *What sort of a value is dignity?*

On the face of it, it is not at all clear what is supposed to tie the three thoughts that compose dignity together: Even if we accept the posited radical discontinuity of value between humans and other things, why should it follow that all humans are of equal worth? And even if we accept both the first two thoughts, why should it follow that the correct way to respond to this equal and higher value of human beings is to 'respect' them and treat them as 'end in themselves'? Why should the higher value of human beings entail a ban on sacrificing one human being to benefit several others?[74]

Insofar as we are attempting to explain dignity in terms of the possession of a special trait, we have two options.[75] Either we posit dignity as a *sui generis* kind of value, which necessitates the three thoughts, or we attempt to explain how dignity is possible in terms of a more familiar sort of value.

The first is Kant's approach. I shall suggest that the second is unfortunately not too promising, as it either must either implicitly presuppose Kant's

---

[74] A sceptic might here push the point that if human beings are fundamentally of equal worth, then it would only be reasonable to do as the consequentialist would and seek to maximise the amount of worth in the world.

[75] When we turn to the possible sceptical justifications, our choice is wider, as we shall see.

theory of value, or simply have no coherent account to give of how we get from 1 and 2 to the third claim, that *respect* is due to human beings.

A positive justification of the dignity of human beings that did not rest on Kant's theory of value would have to have something like the following form:

1. There is a feature or set of features (F) which human beings have, but which nothing non-human has; suitable candidates might be a soul, being made in God's image, or rationality.

2. F accounts for the categorially higher value of human beings.

3. F is present in all human beings to an equal degree; hence human beings are of equal value.

4. Human beings are thus intrinsically valuable insofar as they have F.

The difficulty is in how we are to get from this claim to

Conclusion: Human beings should be respected.

For it in no way follows from the concept of what is intrinsically valuable that it should thereby be respected; we need some further consideration to close the gap. We could attempt to bridge the gap by adding an additional premiss

5. What is intrinsically valuable ought to be respected.

But the problem is that (5) does not seem to me to be true. I cannot think of any other way of bridging the gap, and am thus inclined to think that there is a gap between (4) and the conclusion, of a sort similar to an is-ought gap, but one that is a gap between two different sorts of value claim. It seems to me best therefore that we follow Kant's theory of value.

On Kant's theory of value, dignity is a specific *type* of value, and the three thoughts about dignity are necessitated by the nature of this sort of value.

Dignity, as a type of value, is defined in opposition to another type of value, which Kant calls *price.* Price and dignity are mutually exclusive, and divide the realm of objects between them: hence everything that does not have a dignity has a price.[76] Price and dignity are not commensurable: that is, there is no way of translating between a dignity and a price. Price and dignity are distinguished by their differing underlying logics. Kant expresses these as follows:

> Whatever has a price can be replaced by something else as its equivalent; on the other hand, whatever is above all price, and therefore admits of no equivalent, has a dignity. (*G,* p.434)

Expanding this, we can say: what has a price can, as such, have something else substituted for it; what has a dignity, as such, cannot have something else substituted for it. Therefore, anything that has a price is replaceable; anything that has a dignity is irreplaceable. In the realm of price, everything may be traded or substituted for its equal: the other in which nothing may be traded or substituted for anything else.

What has a dignity must be valued *for its own sake,* and hence it cannot be used just as we please. Kant attempts to explain this by calling something with a dignity an *end in itself.* He gets to this description as follows: he assumes that anything that doesn't have a dignity may be used as a means to whatever ends we may happen to have, and so in trying to capture the thought that there are some things which *may not* legitimately be used as a means to just any ends, but call for a response to them *for their own sake,* he introduces the term end in itself.[77]

---

[76] 'In the kingdom of ends everything has either a price or a dignity.' (*G*, p.434)

[77] It is important to note that the distinction between price and dignity is logically separable from the distinction between intrinsic and extrinsic value. The intrinsic/extrinsic distinction, considered in itself, refers to the location of the value of an object; not to the way we should respond to the object's value. It is conceptually possible that something could have an intrinsic value (has its value in or through itself) but could still legitimately be used as a means to whatever ends we would want to put it to; correlatively it is possible that something might be extrinsically valuable (not having value in itself, but perhaps having valued conferred upon it by us) but yet valued for its own sake (a work of art might be a possible example here). See Korsgaard (1983) on this.

Once we grant this distinction between price and dignity, the unity of dignity and the internal relation of the three thoughts about higher value, equality and respect becomes clear. Whatever has a dignity is *as such* categorially distinct from whatever has a price, and above price. Hence the first thought about the dignity of human beings amounts to no more than this: human beings have a dignity; everything else, a price. The second thought, about the equality of human beings, also drops directly out of the distinction between price and dignity; but the nature of this equality turns out to be different from what one might imagine.

The most obvious way to envisage the equality of human beings is by analogy to an equality of physical magnitude (as when we say two sticks are of equal length); that is, to imagine that the equality of human beings must consist in their possession of an *equal amount* of some special feature. But once we make the distinction between price and dignity, we can no longer explain the equality of human beings in this way. For this way of thinking belongs to the logic of price. To value something as having a dignity is to *refuse* to measure its value against that of other things and to insist that we value it *for its own sake*. Insofar as we wish to attribute a dignity to human beings we cannot talk of a literal equality of value for human beings: we cannot say that human beings are equal because there is some scale of value on which they all score, say, 10.

Rather, the equality of human beings must be understood in a somewhat metaphorical sense. They are said to be *equal,* instead because *each* must be valued for his or her own sake. The equality lies in the *attitude* with which each must treat others and with which each may in turn expect to be treated by them. For Kant, this attitude is *respect:*

> The *respect* that I have for others or that another can require from me (*observantia aliis praestanda*) is therefore recognition of a *dignity* (*dignitas*) in other men, that is, of a worth that has no price, no equivalent for which the object evaluated (*aestimii*) could be exchanged. ... Every man has

a legitimate claim to respect from his fellow men and is *in turn* bound to respect every other.[78]

So the higher worth of human beings, the equality of human beings, and the fact that they ought to be respected all fall out of this basic distinction between price and dignity.

3.4.2 *How could a special trait bestow dignity on its possessor?*

If we now wish to justify the attribution of a dignity to all human beings by reference to a special trait or traits, it is clear that any candidate trait must meet the following constraints:

1. This trait must in fact be possessed by all those to whom morality wishes to attribute a dignity, and not possessed by any being to whom morality does not wish to attribute a dignity.

2. The goodness of this trait must be such that respect is an appropriate response to its possessor.

3. The trait must underwrite equality of treatment.

4. The respect-worthiness of the trait must be compatible with the rest of what we know about human beings.

The point of the first three constraints should be relatively obvious: on the first, given that morality advocates the dignity of *all* human beings, it is clearly deeply problematic if it then argues for this dignity on the basis of a trait that not all human beings in fact have. But it would be equally problematic if the special trait were distributed too widely, and so failed to distinguish human beings from other animals (consider, for example, the capacity to feel pain).

---

[78] *MM,* p.462. See also *MM,* p.435 (quoted in chapter one): man 'possesses a dignity (an absolute inner worth) whereby he exacts respect for himself from all other rational beings in the world. He can measure himself with every other being of this kind and value himself on a footing of equality with them'.

On the second, it would clearly be problematic if we picked as our special trait something like the capacity for instrumental reason, that whilst undeniably a good thing, scarcely seems sufficient to render its possessor incomparable in value to anything that does not possess it.

The third should follow from the first and second, but it is as well to be explicit here: suppose someone claimed that the special trait is the capacity for rational thought. Now this capacity is clearly present in different human beings to differing degrees: why shouldn't this imply a greater value for the more rational? Or we could press the same point from the opposite direction: higher primates are by no means totally lacking in the capacity for rational thought. How do you justify their total exclusion from the realm of dignity?

The fourth should be obvious too. It is no good to refer to an attribute such as possession of an immortal soul, or being made in God's image unless one has some way of demonstrating that this is in fact true.

Kant's theory of value requires that i) everything either has a price or a dignity, and ii) there is a radical discontinuity between price and dignity, and iii) hence there must be some definite answer to the question which category we should place a particular object in.

There are two main problems with this: first, the distinction between price and dignity is supposed to be an all or nothing affair; a thing must be definitely in one camp or in the other. Given the discontinuity between price and dignity, if we attempt to underwrite the distinction with reference to traits that vary *continuously,* our attempts are likely to look arbitrary.

Second, our firmly held moral intuitions rebel against such a simplistic model: to do justice to our moral intuitions we would require a more gradualist model which allows a certain fuzziness around the edges, and especially a sort of value *between* price and dignity.

Think, for example of the question of at what stage a human life should be accorded the value of a dignity. Can we plausibly maintain that there is a certain time during which a human life has merely a price, and then suddenly a threshold is crossed and that life has a dignity? Unless we believe in supernatural claims such as that there is a specific moment of *ensoulment,* then it is difficult to see how we could justify the idea of crossing over from price to dignity in this way: for all we seem to see is continuous development, and hence a gradualist account of value would seem to be more appropriate.

3.4.3 *Are there any special traits that could at all plausibly fit this bill?*

Kant thinks that it is our *humanity* (which the same as our rational nature and our ability to legislate universal law) which is the special trait that gives us our dignity:

> [T]he dignity of humanity consists just in its capacity to legislate universal law, though with the condition of humanity's being at the same time subject to the very same legislation. (*G,* p.440)

He constructs an explicit argument (which he thinks holds a priori) to justify why we should believe in the dignity of all rational creatures:

> If then there is to be a supreme practical principle and, as far as the human will is concerned, a categorical imperative, then it must be such that from the conception of what is necessarily an end for everyone because this end is an end in itself it constitutes an objective principle of the will and can serve as a practical law. The ground of such a principle is: rational nature exists as an end in itself. In this way man necessarily thinks of his own existence; thus far it is a subjective principle of human actions. But in this way also does every other rational being think of his existence on the same rational ground that holds for me; hence it is at the same time an objective principle, from which, as a supreme practical ground, all laws of the will must be able to be derived. (*G,* pp.428-9)

Admittedly, it is slightly unclear how this argument is supposed to work: but for our purposes we can content ourselves with looking at the underlying idea that Kant is attempting to express.[79] This is the thought that it is the *same trait*, 'humanity' which is responsible *both* for the dignity that each person has, *and* for their ability to respond to the dignity of others.[80]

However, this account falls foul of the first constraint we canvassed. Kant is committed to the claim that *only* those things that are capable of respecting dignity themselves have dignity; but if we sincerely carried this thesis through then we would not longer be able to hold on to many of the moral beliefs we feel most sure about; for example that infants who have not yet developed the capacity for respect, and those who through senility have lost it, are still themselves to be treated as if they had a dignity and not just a price.

We can explain this point more fully if we make a distinction between what I shall call the *subjects* and the *objects* of dignity. We can define them as follows: a *subject* of dignity is a being who is capable of *appreciating* dignity in other things; an *object* of dignity is a thing which should be appreciated as having a dignity. Kant's explanation of the dignity of human beings, could be re-expressed in this language by saying that all subjects of dignity are objects of dignity, and they are objects of dignity *just because* they are subjects of dignity.

Kant's further assumption could be expressed as follows: the only objects of dignity are those that are simultaneously subjects of dignity: that is that there is nothing is worthy of being appraised as having a dignity which cannot itself appraise other things as having a dignity. It is this further assumption that is dubious: there would appear to be things such as infants

---

[79] Korsgaard attempts to rework this argument: we shall examine her attempts to do so in the next chapter.

[80] Velleman helpfully elucidates as follows: 'people have a capacity whose value we appreciate by respecting them; and that capacity, at its utmost, is *their* capacity for respect.' (Velleman 1999, p.365)

that we ordinarily take to be objects of dignity without their being subjects of dignity.

But other potential special traits fare no better either: it is extremely difficult to suggest a trait that would satisfy even ourselves as defenders of morality, let alone a more suspicious mind such as Nietzsche's. The basic problem is that the only features that could at all plausibly meet the four constraints focus on elements of human subjectivity: for example, the possession of a unique point of view, the capacity to find things valuable, the capacity for autonomy, or being rational, or the capacity to be a moral agent, perhaps our capacity for repentance. But for each proposed trait, there are categories of human being whom the tradition of morality has always deemed worthy of respect (infants, the mentally defective, for example) who cannot be said with any degree of plausibility to share in the relevant feature or features.

This leaves us in an unfortunate quandary: we have three options available to us. All are unpalatable. The first would be to bite the bullet and come out and say that those who do not have the special feature should not be accorded a dignity: so that infants, for example, would be excluded from the full protection of morality. The second would be to attempt to weaken one's specification of the special feature so that it, for example, required only the *potential* to exercise it at some later date. The third would be to claim that even though *strictly speaking* such persons do not make the grade, there are good reasons for treating them as if they did (we claim, for example, that they are members of a species which in normal circumstances develops the traits which are sufficient for dignity, and so should be treated as if they had a dignity.)

Each approach has its defects. The first risks squandering the moral insights that morality is attempting to safeguard; the second risks weakening the special feature to the extent that possession of it no longer

seems to render appropriate the affirmation we associate with dignity. The third, absent some rather better arguments than it is usually presented with, is just ad hoc.

3.5 *Conclusion: the necessity of a sceptical justification*

The difficulties involved in attempting to give a positive justification thus lead me to think that we are better off going down the route of a sceptical justification. If we pursue a sceptical justification, we will still be able to explain how morality makes life worth living through the concept of the dignity of human beings, but we will not have to give an account of what special trait this dignity consists in.

There are a variety ways in which we might work with a sceptical justification of the dignity of human beings. In the remainder of this thesis I shall consider three.

The first is Christine Korsgaard's reworking of Kant's arguments; the second Charles Taylor's position and the third my own. Korsgaard and Taylor both put forward transcendental arguments which aim to show that believing in the dignity of all human beings is *necessary* for us. Both attempt to argue from the structure of the self to the idea of the dignity of human beings, and hence to the validity of morality.

Korsgaard holds that it is necessary if one is to avoid nihilism that one accords oneself a dignity, and that according oneself a dignity rationally requires one to accord all other value-conferring beings one as well.[81] Charles Taylor holds that according all others a dignity is mandated by the unavoidable attempt to make the best sense of the lives we are trying to lead.[82]

Where these accounts appeal to the transcendental conditions of agency, my account appeals to empirical psychology. It takes belief in the dignity

---

[81] See chapter four.
[82] See chapter five.

of all human beings to be strictly optional: rather than claiming that it is *necessary* to believe in the dignity of all human beings, it admits the contingency of this belief. But, it argues, there are good reasons to live according to this belief nonetheless.

Let us begin by examining Korsgaard's arguments.

C h a p t e r   F o u r

Korsgaard's Argument for the Normativity of Morality

4.1 *Introduction*

In *The Sources of Normativity,* Christine Korsgaard attempts to modernise the Kantian argument for the universal dignity of human beings we looked at in the previous chapter.[83] In so doing to she attempts to use Kantian ideas to construct a sceptical justification for the dignity of human beings by showing that human selfhood implies valuing both one's own humanity (capacity for rational choice) and the humanity of all other human beings.

Here I analyse her approach: by showing up some of its limitations we will take a first step toward constructing a sound argument for morality. Thus, my interest is in the possibility of the *sort* of strategy she employs demonstrating the truth of morality. So in our discussion we are less interested in the *particular* argument that Korsgaard gives, than in the question of whether *any* argument which starts from the reasons implied by self-consciousness could possibly establish an affirmative account of morality. Thus, whenever her argument seems dubious, or to be missing some intermediate premisses that would be necessary to make it work, I attempt to supply them. If Korsgaard's argument, read as sympathetically as possible, and as an exemplar of a whole strategy of argumentation, fails, then this whole approach is wrong, and we must try another.

---

[83] Korsgaard 1996b: in this chapter, page references are to this work unless otherwise stated.

I begin by laying out what I take the structure of the argument to be: this will serve both as a heuristic to the understanding of Korsgaard's position and as a map of my discussion.

Korsgaard's argument aims to show how, by an unbroken chain of implications, our self-consciousness implies that we must value our capacity for rational endorsement, and thereby the capacity for rational endorsement of all human beings.

To make the discussion more manageable, I have separated the argument into several overlapping sub-arguments. The first argument aims to demonstrate that our self-consciousness imports a reflective structure to our consciousness, which implies that we must think of ourselves as acting for *reasons*.[84] The second argument aims to show that the reflective structure of our consciousness implies that we must decide how to act *as if* we were free. (Or, as Kant put it, that we 'act under the idea of freedom'.) It further aims to show that acting under the idea of freedom implies that we act on universal principles. The third argument aims to buttress this conclusion: it argues that we must think of ourselves as aiming to act on universal principles, because the will is a cause; the concept of a cause brings with it the notion of lawlikeness; and hence the free will must will in accordance with lawlike universal principles.

Between them, the first three arguments aim to establish the claim that the reflective structure of our consciousness implies that we act on universal principles. I argue that this claim is not established: all that is established is the weaker claim that the reflective structure of our consciousness implies that we act must in ways that we endorse, and so-acting need not imply acting in accordance with universal principles.

---

[84] As Korsgaard presents it, it is difficult to tell how this argument is supposed to work. I attempt to elucidate it on the assumption that it attempts to show that being a person presupposes that one acts for reasons.

The next five arguments aim to add the content required for morality to the bare idea of acting on universal principles (for given what Korsgaard thinks she has established with the first three arguments, a consistent rational egoist would count as acting on universal principles).

The fourth argument aims to show that a reflectively self-conscious agent must have a self-conception, which she chooses *from,* and which is a description (or set of descriptions) under which the agent values herself. Korsgaard calls this set of descriptions a *practical identity.*

The fifth argument aims to show that nothing is a genuine reason for an agent unless it is implied by a *practical identity* that the agent has, and that obligations spring from what is forbidden by a given practical identity.

The sixth argument is largely my own interpolation: it aims to make clear what would be involved in successfully reconstructing morality the basis of the idea of practical identity. If, following Korsgaard, all obligations spring from practical identities, then if there are to be universal obligations for human beings as such, these obligations must follow from a practical identity that each human being has. But there are two further constraints: given that the claims of morality are supposed to trump claims provided by all other practical identities (say as a football coach, or as a member of the Corleone family), then we must be able to demonstrate that the practical identity appealed to by morality can legitimately claim precedence over all other practical identities. Last, even if an obligation were universal and trumped all others*,* it still would not be capable of giving us morality unless it necessitated us acting on reasons that are public and shareable.[85]

The seventh argument is Korsgaard's attempt to demonstrate that there is a practical identity that is implied by any other practical identity we have, which is a valuation of ourselves *as* reflective choosers, which she calls

---

[85] We could for example have an argument which showed that each person had a practical identity as a rational egoist, and that this practical identity should be put before all others: in this case, despite the fact that we had universal obligations that trumped all others, we would be no closer to the validity of moral obligation.

(following Kant) our *humanity.* There is a weak and a strong reading of this argument: the strong reading would say that humanity is the *source* of all value (and of the value of all practical identities) and hence without it there would be no value; the weak reading would say that the value of humanity is *implied* by having any other practical identity.

We can see from the sixth argument that the total structure of Korsgaard's argument *requires* the strong reading, for the even if the weak reading is true, it gives us no reason why our practical identity as human should be able to claim precedence over any other practical identity. However, the claims about value implied by the strong reading are not true. Further, there are good reasons for thinking that even the weak reading is false.

The chain of argument thus breaks down irreparably at this point.

The eight argument aims to show that our deep social nature commits us to the necessity of conceiving reasons as essentially *public* (agent neutral), on the following grounds: we can think and reason together; therefore reasons cannot be inescapably private; so we must think of reasons as public. This argument is invalid, but this matters little given that the chain of reasoning has already broken down irreparably. But it has the great virtue that it points us in the right direction: towards our social nature in a way that is taken up both by Charles Taylor and by my pragmatist account in part three.

4.2 *From self-consciousness to reason governed action*

The first argument is by far the most important. It aims to demonstrate that a human being is a being that acts for reasons. The remaining arguments (two to eight) are just a working out of what is implied by this idea of a being that acts for reasons.

Korsgaard offers the following as her version of the first argument:

The human mind is self-conscious... in the sense that it is essentially reflective. ... A lower animal's attention is fixed on the world. Its perceptions are its beliefs and its desires are its will. It is engaged in conscious activities, but it is not conscious *of* them. That is, they are not objects of its attention. But we human animals turn our attention on to our perceptions and desires themselves, on to our own mental activities, and we are conscious *of* them. That is why we can think *about* them.

And this sets us a problem no other animal has. It is the problem of the normative. For our capacity to turn our attention on to our own mental activities is also a capacity to distance ourselves from them, and to call them into question. ... I desire and I find myself with a powerful impulse to act. But I back up and bring that impulse into view and then I have a certain distance. Now the impulse doesn't dominate me and now I have a problem. Shall I act? Is this desire really a *reason* to act? The reflective mind cannot settle for perception and desire, not just as such. It needs a *reason.* Otherwise, at least as long as it reflects, it cannot commit itself or go forward. (pp. 92-3)

This passage purports to be offering an argument from a premiss (the claim that the human mind is reflectively self-conscious) to a conclusion (that this forces us to act for reasons), via a middle term (the reflective endorsement of a desire).

However, if we read it in this way it is somewhat unsatisfying, for Korsgaard does not appear to give any sort of argument to get to the middle term: all we have is the claim that 'our capacity to turn our attention on to our own mental activities is also a capacity to distance ourselves from them, and to call them into question', which appears to be an assertion rather than an argument. Further, the move from the middle term to the conclusion seems to rest on a stipulation: for Korsgaard defines a reason *in terms of* reflective endorsement.[86] The move from the middle term (reflective endorsement) to the conclusion (reason governed action) therefore follows by definition.

---

[86] 'We need reasons because our impulses must be able to withstand reflective scrutiny. We have reasons if they do. The normative word 'reason' refers to a kind of reflective success.' (p.93)

We have an argument, then, in which two moves are made: one is a mere assertion, the other a stipulation. It is hardly surprising that it seems unsatisfying.

But Korsgaard gives a hint about where we should look to fill out the argument.[87] She claims there that her account of a person has an obvious affinity with that put forward by Harry Frankfurt,[88] and she implies that human consciousness is *essentially* reflective in a way that Frankfurt doesn't quite realize.

I suggest that this allows us to reconstruct the presuppositions of her argumentative strategy in as follows:

1) Morality is concerned with the obligations owed by *persons*.

2) It follows that the *concept* of a person plays the role of a fixed point in the argument.

3) Therefore, if we can demonstrate that what it is to be a person involves a commitment to acting for reasons then we will have done enough.[89]

I begin by giving Frankfurt's conception of the person: then I shall try to explain why Korsgaard thinks i) it is in fact incorrect as to what is distinctive about persons, and ii) once we realize what is distinctive about persons, it requires that we rethink personhood as essentially active and reflective, in a way that is ruled out by the framework with which Frankfurt is working. We will then show that acting for reasons is presupposed by this revised conception of a person.

---

[87] p. 99.

[88] Frankfurt (1971)

[89] There is an obvious risk with this strategy, analogous to the one we saw with attempted positive justifications of the dignity of human beings, namely that our definition of person is actually unhelpful for moral thinking because it draws the boundaries in the wrong place or makes firm distinctions where there are in fact none. Bernard Williams draws this worry out well: "The category of person, though a lot has been made of it in some moral philosophy is a poor foundation for ethical thought, in particular because it looks like a sortal or classificatory notion while it in fact signals characteristics that almost all come in degrees – responsibility, self-consciousness, capacity for reflection and so one. It thus makes it seem as if we were dealing with a certain class or type of creature, when in fact we are vaguely considering those human beings who pass some mark on a scale." (Williams 1985, p.114)

Frankfurt's conception of a person relies on a distinction between first and second-order desires. First-order desires are 'simply desires to do or not do one thing or another', (1971, p.12) whilst second-order desires are desires about the sort of first-order desires we would like to have. Whilst animals can have first-order desires, they cannot have second-order desires: these are the sole preserve of human beings.[90]

Frankfurt then gives a definition of a *will:*

> To identify an agent's will is either to identify the desire (or desires) by which he is motivated in some action he performs or to identify the desire (or desires) by which he will or would be motivated when he acts. An agent's will, then, is identical with one or more of his first-order desires. But the notion of the will, as I am employing it, is not coextensive with the notion of first-order desires. It is not the notion of something that merely inclines an agent in some degree to act in a certain way. Rather it is the notion of an *effective* desire – one that moves (or will or would move) a person all the way to action. (1971, p. 14)

When we apply the distinction between first and second-order desires to this notion of will, we obtain the notion of first and second-order volitions. Thus, an entity has a second-order volition when 'he wants a certain desire to be his will' (1971, p.16). It is, according to Frankfurt, the capacity for *second-order volition* that is the mark of a person.

This completes the necessary distinctions. Frankfurt then attempts to show why these definitions capture the concept of a person by applying them to a concrete example. He asks us to consider two agents, both of whom are drug addicts, the first being a person and thus capable of second-order volition, the second a *wanton* (Frankfurt's technical term for an agent who has desires but who is incapable of second order volition.)

---

[90] 'Besides wanting and choosing and being moved *to do* this or that, men may also want to have (or not to have) certain desires and motives. They are capable of wanting to be different, in their preferences and purposes, from what they are. Many animals appear to have the capacity for what I shall call 'first-order desires' or 'desires of the first order,' which are simply desires to do or not do one thing or another. No animal other than man, however, appears to have the capacity for reflective self-evaluation that is manifested in the formation of second-order desires.' (1971, p.12)

The wanton can be no more than a space in which various desires play themselves out: while 'he' can be rational in a sense, for example, by adopting the best means given the desires he has, we must remember that there is nothing to 'him' apart from these first-order desires:

> He does not prefer that one of his conflicting desires should be paramount over the other; he does not prefer that one first-order desire rather than the other should constitute his will. It would be misleading to say that he is neutral as to the conflict between his desires, since this would suggest that he regards them as equally acceptable. Since he has no identity apart from his first-order desires, it is true neither that he prefers one to the other nor that he prefers not to take sides. (1971, p.18)

Certainly, it seems that we would not call such an entity a person.

He contrasts this with the *person,* to whom *it does* make a difference which of his first-order desires wins out:

> It makes a difference to the unwilling addict, who is a person, which of his conflicting first-order desires wins out. Both desires are his, to be sure; and whether he finally takes the drug or finally succeeds in refraining from taking it, he acts to satisfy what is in a literal sense his own desire. In either case he does something he himself wants to do, and he does it not because of some external influence whose aim happens to coincide with his own but because of his desire to do it. The unwilling addict identifies himself, however, through the formation of a second-order volition, with one rather than with the other of his conflicting first-order desires. He makes one of them more truly his own and, in so doing, he withdraws himself from the other. (1971, p.18)

It is notable that despite Frankfurt's invocation of second-order desires and volitions, his view of persons is basically Humean: where a straightforward Humean view would view the person as a bundle of desires, Frankfurt complicates the picture by arguing that there are desires about desires. However, according to Korsgaard, the same Kantian worry about what *unifies* the Humean self applies.

While Korsgaard would agree with Frankfurt that the ability to form second-order volitions in his sense is a *necessary* condition for personhood, she does not think that it is a sufficient condition. It leaves us with a conception of the person that is uninvitingly passive: what she feels this account leaves out is the element of activity. On her account, what makes a person a person is a conception of themselves *as a unity (an 'I'), through which they deliberate and act on the basis of reasons.*

Korsgaard suggests that if we look at what deliberation is like we will see that it presupposes a unity, an 'I', a standpoint *from which* deliberation takes place:

> It may be that what actually happens when you make a choice is that the strongest of your conflicting desires wins. But that is not the way you think of it when you deliberate. When you deliberate, it is as if there were something over and above all your desires, something that is *you,* and that *chooses* which one to act on. The idea that you choose among your conflicting desires, rather than just waiting to see which one wins, suggests that you have reasons for or against acting on them. And it is these reasons, rather than the desires themselves, which are expressive of your will. (1989a, p.370)

Subjectivity of this sort (considering oneself as an 'I' who chooses) is essential to being a person. Moreover, it enjoys an explanatory priority over Frankfurt's favoured conception of second-order volitions. While we can explain the capacity for second-order volitions if we take the basic element in a person to be an 'I' who chooses, we cannot explain the subjectivity – the 'I' in terms of a capacity for second-order volition.

Given that a person is at least an 'I' who chooses, if we work out the conditions under which it is possible to have an 'I' who chooses, we know that we can presuppose this of ourselves given that we are persons. This will take a form something like the following. Inevitably it anticipates some of the points covered by arguments two to eight, as it does not try to show how we must act for reasons from non-normative premisses, but

instead attempts to make this thesis plausible by laying out a structure of how things seem to fit together. Parts of this structure are also involved in arguments two to eight.

From an I who chooses *to reason governed action*

1. Personhood presupposes an 'I' or a 'self'.

2. This 'I' unifies the human being into a single entity.

3. This unification is both *reflexive* and *active.*

3a) It is *reflexive* because the unity consists in the conception that the self has of itself. (i.e. the unity isn't grounded in anything further than the self's conception of itself).

3b) It is *active* because the unity of the self depends on the self being able to actively maintain a conception of itself. (Thus it can be shattered if for any reason the self is no longer able to maintain a conception of itself as a unified being.)

4. A person *chooses* what to do.

5. When a person freely chooses what to do she does so in a way that is expressive of herself.

6. Therefore a person's conception of herself does double duty: it both holds her together as a unified agent and acts as a standard against which she judges what to do.

7. This self-conception must be one that the agent reflectively endorses: otherwise it could not fulfil its role, which is to demarcate what the agent considers herself to be.

8. Given that i) the self's conception of itself implies reflective endorsement of itself, and ii) the capacity for choice is not external to the

self, choosing in accordance with the capacity for choice must involve choosing what one reflectively endorses.

9. Given that one is choosing what one reflectively endorses, one is choosing for reasons.

10. Persons act for reasons.

This argument allows us to help ourselves to the conclusion that normative judgment is already implied in our conception of ourselves as persons, and that hence normativity is not something we need discover in the world, but something presupposed in what we are. This is an important result if we want to give (as Korsgaard does) an account of normativity that allows us to see it as continuous with the world of natural facts. But as this is not our focus here, I merely mention it and pass on.

4.3 *From acting for reasons to universal principles*

Argument two starts from the assumption that persons act for reasons and attempts to demonstrate that reasons will have the form of universal principles. Argument two is taken straight from Kant: as it seems to me that he puts the arguments more clearly than Korsgaard, I shall for the most part analyse the relevant passage of the *Grounding of the Metaphysics of Morals:*

> Now we cannot possibly think of a reason that consciously lets itself be directed from outside as regards its judgments; for in that case the subject would ascribe the determination of his faculty of judgment not to his reason, but to an impulse. Reason must regard itself as the author of its principles independently of foreign influences. Therefore as practical reason or as the will of a rational being must reason regard itself as free. (*G,* p.448)

The reading I propose for argument two, on the basis of this passage, is as follows:

1) We start with a blank conception of an agent who acts for reasons.

2) By definition, we know that we can already impute to this agent the capacity for reflectively endorsing her desires.

3) This reflective endorsement cannot come from a source *outside* the agent's capacity for choice, as then it would seem to the agent that she was merely being buffeted by her impulses.

4) The reflective endorsement must then come from *inside* the agent's capacity for choice.

5) The capacity for choice is capable of determining the agent's actions.

6) (From 4) When the agent acts in accordance with what she endorses with her capacity for choice, she must see herself as having the final say on what she does.

7) When an agent acts in accordance with her capacity for choice she must consider herself as free.

8) In a formal sense a free will is a will which makes choices independently of all alien influences.[91]

9) When the will wills freely it must reflectively endorse the action it performs.

10) Hence (by definition) when the will wills freely it must will for a reason.

11) Hence there must be some standard[92] in virtue of which the will endorses the impulses that it does.

12) Therefore, when the will wills freely it must will in accordance with a universal principle.[93]

---

[91] Korsgaard, (1989b, p.162); *G,* p.446.

[92] I intend 'standard' here to be taken in a purely formal sense: in the sense that normativity as such presupposes a standard in virtue of which what is endorse *as* normative is said to be correct. See my proposal for a rewriting of argument three (below) for this.

[93] There is an alternative formulation for this argument, which we could call argument two´, which would go as follows:

  1. Persons are capable of acting for reasons.

The problem with this argument is the move from 11 to 12: while it is true that the idea of reflective endorsement presupposes *some* process in virtue of which certain impulses to act are anointed as *reasons*, it does not seem to follow immediately that this process must consist of deciding from a universal principle.[94] Why couldn't the appropriate model be perceptual, for instance?

Argument three: from the will as a cause to the lawlikeness of reasons

Argument three aims to answer this criticism, and to explain why we must consider ourselves as acting on universal principles. Argument three depends on the following argument of Kant's:

> The will is a kind of causality belonging to living things insofar as they are rational; freedom would be the property of this causality that makes it effective independent of any determination by alien causes. ... The concept of causality involves that of laws according to which something that we call cause must entail something else – namely, the effect. Therefore freedom is certainly not lawless, even though it is not a property of the will in accordance with laws of nature. It must, rather, be a causality in accordance with immutable laws, which, to be sure, is of a special kind; otherwise a free will would be something absurd. ... What else can freedom of the will be but autonomy, i.e., the property that the will has of being a law unto itself? (*G,* pp. 446-7)

As it stands, there are two major problems with this argument. First, it seems to be premised on an equivocation. When we call the will a 'cause', we are using this term in a different way from when we describe a physical

---

2. The capacity to act for reasons presupposes a capacity for judgment ('reason'), which does the reflective endorsing.

3. Reason cannot *consciously* let itself be directed from outside as regards its judgments. It must be able to assent to the judgments it makes.

4. Reason qua reason cannot judge at random (or, 'for no reason').

5. (From 3,4) There must be some standard in virtue of which reason judges which it assents to.

6. This standard will be a *principle* that it lays down for itself.

The move from 5 to 6 is flawed for the same reasons as the move from 11 to 12 in argument two.

[94] As G. A. Cohen puts it, 'The reflective structure of human consciousness may require, as Korsgaard says… I endorse the first-order impulses on which I act. But it does not follow, and it is not true, that the structure of my consciousness requires that I identify myself with some law or principle. … What the reflective structure requires, if anything, is not that I be a law to myself, but that I be in command of myself. And sometimes the commands that I issue will be singular, not universal.' (1996, p.176)

cause: we call the will a cause because it appears to the person whose will it is that their willing is the *first cause* of the action that then ensues. 'Cause' here seems to denote a conviction we have that we are the authors of our actions – in Aristotle's words, that 'a human being originates and fathers his own actions as he fathers his children.'[95] But it remains to be shown that the concept of an apparent *first cause* brings with it the idea of lawlikeness.

Quite apart from this, it is rather difficult to see, without any further argument, why freedom should *require* that an agent act according to laws. Nagel puts this doubt well in his reply to Korsgaard's initial formulation of this argument:

> A neo-Humean regularity theory of causation seems an inappropriate model for free self-determination. If the idea makes sense at all, the free choice of actions which conform to a law is no more nor less a form of *causality* than the free choice of actions which do not... So far as I can see, choosing freely in a law-like pattern is merely a way of mimicking causality; if I always put on my left sock before my right, that does nothing to establish the causality of my will, so why does the categorical imperative do any better? (1996, p.202)

Korsgaard then produces a new argument in an attempt to fill out why Kant's claim is true.

The structure of this argument is far from clear: I shall present the best reading that I can give of it. Korsgaard starts from a certain parallelism of structure that she claims to detect between the concepts of *reason* and *cause:* both allegedly combine a sort of necessitation with a universality claim. Realising that this universality is the very thing that she is supposed to be arguing for she then puts forward the following argument:

1) We start from Hume's claim that since we do not directly perceive power, we cannot perceive individual exercises of power.

---

[95] *Nichomachean Ethics* 11113b18-19.

2) Ergo, *even if* there were such things as single, anomalous instances of one thing *making* another happen, *we* would have no way of distinguishing these from mere temporal sequences of events.

3) 'To regard oneself as an agent is to regard oneself as a cause, as productive of certain actions and their effects. And given the connection between causality and regularity, to do that must be to regard oneself as productive of these actions and effects in some regular way.' (p. 227)

4) 'If I am to constitute *myself* as the cause of an action, then I must be able to distinguish between *my* causing the action and some desire or impulse that is 'in me' causing my body to act. I must be able to see *myself* as something that is distinct from any of my particular, first-order, impulses and motives, as the reflective standpoint in any case requires.' (pp. 227-8)

5) '[I]f *all* my decisions were particular and anomalous, there would be no identifiable difference between *my acting* and *an assortment of first-order impulses being causally effective in or through my body.* And then there would *be* no self – no mind – no me – who is the one who does the act. ... Just as the special relation between cause and effect, the necessitation that makes their relation different from mere temporal sequence, cannot be established in the absence of law or regularity, so the special relation between agent and action, the necessitation that makes that relation different from an event's merely taking place in the agent's body, cannot be established in the absence of at least a claim to law or universality. So I need to will universally in order to see my action as something which *I do.* Nagel misses the point when he says that regularity does nothing to establish the causality of my will. What it does is establish my own ability to see myself as having a will, as having the kind of *self-conscious* causality that *is* a rational will.' (pp. 228-9)

There are two arguments here: first, if we look at (3), then it is clear that there are some remnants of Kant's original argument. But there is also

another, more interesting argument: this one tries to tie willing in a universal manner to the possibility of the will constituting itself as a rational will. This is a better argument, but it founders at the same point that argument two did: we can agree about the key place of normativity in the unity of the self, but it follows that it is only qua normative that the universality is required, and it need not therefore extend as far as universal principles as such.[96]

We need to stand back for a moment and take stock, to see why this should be so. The sort of self-consciousness we have commits us to normative judgment, and combining the idea of normative judgment with the unity of agency presupposed in the deliberative perspective gives us the result that we must will in accordance with some sort of standard which we endorse, if we are to be free. This doesn't actually say that much: we could summarize it as follows – given that we are normative beings, we must will in accordance with our conception of what we normatively endorse. Given that the concept of normativity already involves the notion of standard, it follows that willing freely involves willing in accordance with a standard that we endorse. We haven't got as far as Korsgaard thinks we have, simply because the result comes about because of what is presupposed in the concept of normative judgment, and not due to any external argument about the will as a cause and causes having to operate in accordance with laws.

The difference between the concept of normative judgment in general and the concept of a judgment which operates under laws is as follows: it is right to say that a correct normative judgment (tautologically) must be discursively redeemable as a universally valid claim. However it does not

---

[96] There is an additional problem with this argument, of which Korsgaard herself is aware: even if the argument worked, it could only establish that I must *for the most part* will universally. Even on the argument's own terms, it is false to claim that isolated incidents of anomalous or arbitrary willing would destroy a person's ability to see themselves as having a will. The most that the argument can establish is that one should not act anomalously or arbitrarily *so much* than one undermines one's sense of oneself as having a will.

follow from this that judgment that is normatively correct must be conceived as a following of rules that are already given prior to the exercise of judgment. For example, the particularity of a situation may be such that there is no rule in existence which adequately covers it, and even if there were a rule it would apply to such a small variety of cases that it could not be universalizable in any *useful* sense.[97]

In other cases (possibly an apparent dilemma), moral thought will attempt to *construct* a law that would be appropriate in this and other relevantly similar cases. Here it does not follow that one should will in a lawlike manner; rather one should ask 'what is to be done' – it's just that a *correct* answer to this question will have a universal applicability across all relevantly similar cases. Here, one should not will qua universal, but qua correct response to the problem in hand.

This brings us to the end of the first phase of the argument. In summing up where it has brought us to, Korsgaard admits that it leaves us quite some way from morality. She then makes a distinction that is quite unKantian in inspiration. She suggests we separate i) the conclusion she has just established, which is the requirement to act only on maxims you can will to be laws,[98] which she will call 'the categorical imperative', from ii) what she calls the 'moral law' which is to be the 'law of what Kant calls the Kingdom of Ends, the republic of all rational beings,' which tells us to 'act only on maxims that all rational beings could agree to act together in a workable cooperative system.' (p.99)

---

[97] This is how Aristotle conceives the subject matter of ethics: cf *Nichomachean Ethics* 1137b 14ff. 'All law is universal, but in some areas no universal rule can be correct…. This is also the reason why not everything is guided by law. For on some matters legislation is impossible, and so a decree is needed.'

[98] As we have seen, this holds only in a negative sense, and not in a positive sense: that is, the categorical imperative is a *side effect* of the universality involved in normativity. While a correct result trivially can be willed as a universal law, it does not follow that one must get to the correct result by asking what can be willed universally.

Understood this way, the categorical imperative and the moral law are quite separate entities: the categorical imperative does not set any limits on what may be done except that the agent has to be *consistent* in it. Thus, the rational egoist, who makes it a law that he should always maximise his self-interest, is acting in accordance with the categorical imperative so understood. The moral law, however, is much more substantive, and entails the validity of morality.

4.4 *From the categorical imperative to the moral law*

Arguments four to eight try to make the move from the categorical imperative so understood to the moral law.

Argument four: from self-consciousness to practical identity

Argument four takes place within the space of a paragraph:

> The reflective structure of the mind is a source of 'self-consciousness' because it forces us to have a *conception* of ourselves. As Kant argues, this is a fact about what it is *like* to be reflectively conscious and it does not prove the existence of a metaphysical self... When you deliberate, it is as if there were something over and above all of your desires, something which is *you,* and which chooses which desire to act on. This means that the principle or law by which you determine your actions is one that you regard as being expressive of *yourself.* To identify with such a principle or way of choosing is to be, in St Paul's famous phrase, a law to yourself. (p.100)

Korsgaard does not make it entirely clear how we are supposed to construe this argument. I suggest we see it as the complement to argument two above. Argument two shows that the concept of judgment implies some sort of standard in virtue of which you are judging, which, as we saw, we could in a negative sense say is the form of universality. This argument must try to show that the standard(s) *which* are used in judgment must have some sort of content; that is, the standards must not only be *universal in form,* but *also* must have content.

114

We already know from argument two that an agent is committed to acting in a way that he endorses. This argument tries to elucidate the point that there must be something *in virtue of which* he endorses the things he does. It will do this by a consideration of two obvious facts about judgment: (1) judgment must be *sufficient* to determine what the agent is to do in normal situations; (2) the concept of judgment we employ must be able to support the idea that is closely tied to the concept of agency, that of the agent's *ownership* of his actions.

Both considerations show that it is incoherent to have a concept of judgment that does not presuppose some content *from which* the agent judges.

1) The categorical imperative, as defined by Korsgaard, commits the agent only to acting in a way that they could will as a law. Now, in order for this to represent a normative standard in terms of which the agent is capable of judging himself, there has to be something that would count as transgressing it. That is, it must be possible for the agent to act inconsistently (or as Hegel might have put it, it must be possible for the agent to contradict himself). But inconsistency implies that there is something fixed in the agent, that the agent can be inconsistent with.

Further, imagine that (*per impossible*) there were an agent that had no determinate identity or prior values, which he was capable of being consistent or inconsistent with. If such an agent just had the maxim of willing consistently, then this would not rule out any actions. His faculty of judgment would radically underdetermine what the agent should do in any given situation, and hence would fail in its own terms as a faculty of judgment.

2) Unless the agent did have a determinate identity to choose from it would be impossible for him to have the experience of ownership of his acts. This is because our feeling of ownership of our actions derives in great part

from our feeling that we have acted in a way that is *expressive* of ourselves. For if it were the case that it were the form of universality *per se* that determined a particular action, then there would be nothing in that act that would allow the agent to see it as specifically *his.*

It follows then, from both of these considerations, that we must impute to the agent some conception of himself which provides the content for actual judgment. The concept of a *practical identity* is then introduced to give us a better grip on this content that lies behind the mere universal form. A practical identity is to be understood as 'a description under which you find your life to be worth living and your actions to be worth undertaking.'(p.101)[99]

Argument five: from practical identity to the nature of obligation

Argument five goes as follows:

> Practical identity is a complex matter and for the average person there will be a jumble of such conceptions. You are a human being, a woman or a man, an adherent of a certain religion, a member of a certain ethnic group, a member of a certain profession, someone's lover or friend, and so on. And all of these identities give rise to reasons and obligations. Your reasons express your identity, your nature; your obligations spring from what that identity forbids. (p.101)

This argument is no less compressed than the others that we have looked at: but it also harbours a further perplexity. *Prima facie,* it appears that Korsgaard is now using 'reason' differently from the definition I quoted earlier. For Korsgaard there defined a reason in terms of reflective endorsement, but now it appears that reasons are merely *given* by our practical identities.

---

[99] There is of course a more direct route to the conclusion that argument four offers: in our mapping of the concept of a person, we noticed that having an 'I' that chooses already presupposes that the I unifies itself through a conception of itself.

Suppose an agent grows up in England: presumably it must be part of his practical identity to be an Englishman. But if the practical identity of an Englishman *just gives the agent reasons* (perhaps to enjoy cricket and drink tea) then it appears that the agent will have reasons whether or not he reflectively endorses them; or, if we try to hang on to the conception of reflective endorsement, and put it alongside that of practical identity, then it will seem that the practical identity seems insufficiently reflective – it seems to do no more than to *rubberstamp* desires.

Korsgaard's answer to this is that practical identities themselves are within the scope of reflective endorsement, and that it is only insofar as we endorse them that they are capable of providing us with reasons. A practical identity is not to be treated as a mere given – something, for example, that the agent *happens* to be born into: rather, the practical identity is itself subject to reflective endorsement.[100] This is exactly what we would expect given our preliminary examination of the concept of a person, where we saw that a person must actively maintain the conception they have of themselves.

The overall picture is this: particular judgment, and thereby reflective endorsement of desires, presupposes a background identity that is treated as more or less fixed for the purposes of this particular judgment. But this background identity can perform the role of being a standard for judgment only if it is itself reflectively endorsed by the agent. Normativity therefore permeates the structure of the self: both in its particular judgments and in the identity that makes those judgments possible.

---

[100] See p.123 for Korsgaard's recognition of this: in her terms, this entails that we have a normative conception of our identity.

Argument six: from obligation to the presuppositions of the moral law

Argument six is not one that Korsgaard explicitly formulates: rather, it is my attempt to elucidate what is required for Korsgaard's arguments seven and eight to be successful.

We already know from argument five that all obligations spring from practical identities. It remains to work out how obligations of the right sort to ground the moral law are possible.

First, the relevant obligations must be *universal:* that is, they must apply to each and every human being. Now, taking the premiss that all obligations spring from practical identities, this leaves us with two possibilities: a) there is a practical identity such that it is *necessary* for human beings and that is suitable for grounding the moral law; b) every human being is such that they have at least one practical identity from which it is possible to derive an obligation which is suitable to ground the moral law. In fact, as we shall see from argument seven, Korsgaard holds *both* a) and b).

Second, the relevant practical identity must also enjoy some sort of a priority over any other competing practical identity: it will not be sufficient for morality if the argument provides only an identity that can be defeated by claims coming from one's identity as a member of the Corleone family, or as a rational egoist.[101]

Third, the relevant obligations must also be *public.* 'Public' is here defined in contradistinction to 'private'. Think of the rational egoist: he will be concerned to further his own happiness, but the happiness of others will not matter to him in the same way. If he follows his creed strictly, he will take the happiness of others into account only as far as it enables him to

---

[101] It is worth remembering here that Kant defines the evil person as one who places his own happiness above the moral law: he does not *deny* his practical identity as a moral person, but *subordinates it* to the claims of his own happiness. Thus, if Korsgaard's argument were to establish the necessity of having a practical identity *as a moral being,* but not that this identity must claim precedence, she would have established no more than that we must be at least evil (and not, for example, devilish) in Kant's sense.

maximise his own. For the rational egoist, all reasons are private: they relate back only to him and his life. But private reasons would be insufficient to ground the moral law:

> If reasons *were* essentially private, consistency would not force me to take your reasons into account. And even if it did, it would do it in the wrong way. It would show that I have an obligation *to myself* to treat you in ways that respect the value which I place on you. It would show that I have duties with respect to you, about you, but not that there are things I owe *to* you. (p.134)

Thus the validity of the moral law requires that we show that reasons are public.

Argument seven attempts to establish the first and second conditions, and argument eight the third.

Argument seven: from possession of any practical identity to valuing your humanity

Argument seven is the most difficult to construe. Just before the following passage, Korsgaard has been talking of the possibility of revising one's practical identities in the light of reflection; suggesting 'most of the self-conceptions which govern us are contingent' (p.120):

> What is not contingent is that you must be governed by *some* conception of your practical identity. For unless you are committed to some conception of your practical identity, you will lose your grip on yourself as having any reason to do one thing rather than another – and with it, your grip on yourself as having any reason to live and act at all. But *this* reason for conforming to your particular practical identities is not a reason that *springs from* one of those particular practical identities. It is a reason that springs from your humanity itself, from your identity simply as *a human being,* a reflective animal who needs reasons to act and live. And so it is a reason you have only if you treat your humanity as a practical, normative, form of identity, that is, if you value yourself as a human being. (pp. 120-1)

119

We can agree that having a practical identity is non-optional for human beings: but our confidence in this claim, *pace* Korsgaard, derives from our psychological knowledge of what it is to be a normal human being, rather than from reflection on what it is to have a self-conscious being who acts for reasons.[102] The next part of the passage is problematic, however: there is a jump between the idea of all human beings needing *a* practical identity, and the idea of each human being having a reason for conforming to the particular practical identities they in fact have.

If this is supposed to represent an inference of some sort, then it is invalid: if the argument is supposed to go from the (unstated) premiss that an agent must have a reason to maintain those conditions that make her agency possible, then all that follows is that the agent has a reason that does not spring from any of her current practical identities to maintain *some* practical identity, not this *particular* one.

Korsgaard then attempts to elucidate the argument as follows:

> It is because we are human that we must act in the light of practical conceptions of our identity, and this means that their importance is partly derived from the importance of being human. We must conform to them not merely for the reasons that caused us to adopt them in the first place, but because being human requires it. You may give up one of your contingent practical roles. But so long as you remain committed to a role, and yet fail to meet the obligations it generates, you fail yourself as a human being, as well as failing in that role. And if you fail in all of your roles – if you live at random, without integrity or principle, then you will lose your grip on yourself as one who has any reason to live and to act at all. (p.121)

---

[102] As Gibbard puts it, 'A connoisseur may act form pride in his connoisseurship, from his self-image as a connoisseur, but children seek out pictures from straight liking, and an adult might visit the Art Institute in the same spirit. Not all motivations focus on maintaining a view of oneself as valuable, and a reflective agent might conceivably have no self-focused, esteem-driven motives at all. Without these motives, he would, perhaps, be a normal member of our species – but with Korsgaard, as I read her, we're examining the sheer logic of agency'. (1999, p.152) Note that Taylor's transcendental argument, which we shall look at in the next chapter only aims to show that a particular structure is presupposed by *undamaged human personhood* and thus evades this complaint.

I would call this the argument from integrity: reconstructed, it goes as follows:

1. A reflective agent, by committing herself to autonomy (being a law to herself), thereby also commits herself to integrity (being true to the practical identities she has) in her action.

2. Given that autonomy is a transcendental presupposition of reflective agency, so must integrity be.

3. When I act in a way that lacks integrity I fail *qua* agent with *that* role, but also *qua* agent *simpliciter.*

4. That is, I have failed to live up to my own conception of what I should do.

5. We can only fail *qua* agent if we presuppose an identity *qua* agent which brings with it the obligation to integrity.

6. This identity, as with all others, must be a description under which we value ourselves.

7. Unlike all other identities it is non-optional since it is a transcendental condition for valuing anything at all.

8. This identity involves our valuing ourselves *qua reflective agents* (Korsgaard's claim that it is *qua* human is a little bit misleading here).[103]

9. Sincerely valuing anything brings with it a commitment to integrity – to act on the value when required. To fail to act on the values you endorse is to lack integrity: it matters to you when you lack integrity. This is because what lies behind your concern with integrity is a value that you place on yourself *qua reflective agent:* to lack integrity is to fail yourself *qua reflective agent.*

---

[103] Cf the discussion in the earlier paper Korsgaard 1986.

She then follows up with another, more radical argument, which is designed to show that our humanity is the *source* of all value:

> Most of the time, our reasons for action spring from our more contingent and local identities. But part of the normative force of those reasons springs from the value we place on ourselves as human beings who need such identities. In this way all value depends on the value of humanity; other forms of practical identity matter in part because humanity requires them. Moral identity and the obligations it carries with it are therefore inescapable and pervasive. Not every form of practical identity is contingent or relative after all: moral identity is necessary. (pp.121-2)

> This is just a fancy new model of an argument that first appeared in a much simpler form, Kant's argument for his Formula of Humanity. The form of relativism with which Kant began was the most elementary one we encounter – the relativism of value to human desires and interests ... Kant saw that we take things to be important because they are important to us – and he concluded that we must therefore take ourselves to be important. In this way, the value of humanity itself is implicit in every human choice. If complete normative scepticism is to be avoided – if there is such a thing as a reason for action – then humanity, as the source of all reasons and values, must be valued for its own sake. (p.122)

If this argument works, it is supposed to establish that 'our identity as moral beings – as people who value themselves as human beings – stands behind our more particular practical identities.'

The first thing we need to clarify is what exactly is meant by 'humanity' and what it is to value one's humanity. It seems to me that there is a slippage between the idea of 'humanity' that Korsgaard has argued for, namely one's humanity as capacity for reflective choice, and what she attempts to do with the concept of humanity, which involves a rather thicker notion of humanity. It is unclear that any constraints on action follow from valuing one's humanity in this thin sense.[104] But yet if

---

[104] As Gibbard puts it, 'If valuing my humanity is taking pride in being a reflective chooser, how does that constrain what I do, what I reflectively choose?' (1999, p.156)

Korsgaard wants to give humanity a thicker sense, valuing one's humanity will not follow from the very idea of a self-consciousness.

The second thing we need to clarify with these arguments is the status that Korsgaard is claiming for the value of humanity: there seems to be a degree of sliding about between the relatively weak claim that having any practical identity presupposes attributing a value to our humanity, the slightly stronger claim that in addition our humanity is a partial source of other values, and the very strong claim that our humanity is the *source* of all value.

I suspect that this uncertainty may in part be because Korsgaard realises she requires the very strong claim, but does not have an adequate argument to back it up. She requires this strong claim because if our humanity provides us with just one practical identity amongst others, then there will be no reason to believe that the reasons it provides need be taken to be any more authoritative than those provided by any other practical identity.

However, the strong claim seems to me to be deeply flawed because it ignores two major ways in which value is created in human life without depending directly on the agent's capacity for reflective endorsement: through *activity* and through *recognition.*

First, people often form practical identities *in response to* their finding the activities implied by those identities valuable and enjoyable. Thus, for example, most professional chess players do not start by reflectively endorsing the practical identity of a chess player: rather they begin to play chess, find they are good at it, that they enjoy it, that it begins to interest them more than anything else, *and only then* do they begin to think of themselves as chess players, and to endorse this practical identity. In such cases, it is false to claim that the resultant identity is valuable to that person because they endorse it. This gets things the wrong way round: it would be better to say that they endorse the identity because they find the activity

valuable. Their reflective endorsement is not the source of the value, but a recognition of the value that they had already been attracted by.

Second, it neglects the way in which people can find their lives valuable through being found valuable by others. The way that most of us, if we are lucky, come to have a sense of self-worth if through the internalisation of the love our parents show us. One takes the way one is treated by the other and takes it as an indication of one's worth. The sense of oneself as valuable is *constructed* by love and it is not dependent on one's capacity for rational endorsement.[105]

It thus seems implausible to hold that reflective endorsement is the source of all value. But there are reasons for doubting the weak claim, that having any practical identity presupposes valuing one's capacity for reflective choice, as well. For it is quite conceivable that one could have a particular practical identity without thereby endorsing one's value as a reflective chooser. Let's say I've always considered myself stupid, useless at everything that involves thinking; but I discover that I am a brilliant footballer. I find both immense satisfaction in playing football, and receive extensive validation from others as a result of my footballing excellence. Why should I agree that the value of my identity as a footballer depends on my endorsement of this practical identity and hence on valuing my capacity for rational endorsement in general?

I could doubt that the value of my identity depends on my endorsement as follows: as in the case of the chess player, I might have just stumbled into playing football. What makes it valuable to me is the fact that I'm good at it; that I love doing it; and that (for once) it makes people admire me. In this case the value of the identity comes from the two other sources I mentioned – activity and recognition, and can bypass reflective endorsement altogether.

---

[105] I discuss such relationships of recognition at far greater length in chapter six.

Further, as I suggested I might even have reasons for distrusting and disliking those elements of life that have to do with thinking and with rational weighing of alternatives, and actually wish to be free from all this standing back from life and asking myself if I really endorse the identity I find myself with. Such moves may merely lead me into doubt and uncertainty; make me feel stupid and inadequate; worthless compared to those who are much cleverer than I am. Mightn't I wish that I could be free from this extra layer of consciousness, living for glory and the activity that I love?[106]

To sum up: Korsgaard wants to argue that valuing anything in the world implies that we value ourselves; and that this involves valuing our capacity for rational endorsement. But this argument does not work, as we can value our own lives without thinking that the *source* of this value is our capacity for reflective choice: we could conceivably value our own lives without valuing our capacity for reflective choice at all.

The chain of Korsgaard's argument therefore breaks down at this point.

Argument eight: From our deep social nature to the necessity of conceiving reasons as essentially public (agent neutral)

Where argument seven was supposed to establish the universality of the practical identity of humanity, argument eight is supposed to show that the value we place on our own humanity must be construed in an agent-neutral way.

Korsgaard says that there are only two ways in which we could establish the publicity of reasons: either by asserting the truth of what she calls substantive moral realism (i.e. the view that there are some moral facts or values that are a part of the furniture of the universe, which make our

---

[106] Interestingly, Buddhism too would see something problematic in Korsgaard's idea of reflective consciousness, though for different reasons: it sees it as something that has to be left behind on the road to enlightenment. For a good, philosophically astute introduction to Zen Buddhism, see Sekida 1976.

moral claims justified), or an alternative, which starts from anti-realist premisses:

> The other way retains one element of the picture I began with. The public character of reasons is indeed created by the reciprocal exchange, the sharing, of the reasons of individuals. But it acknowledges the point made by the criticisms made above. If these reasons really were essentially private, it would be impossible to exchange or to share them. So their privacy must be incidental or ephemeral; they must be inherently shareable. We might call this view 'publicity as shareability'. I take this to be equivalent to another thesis, namely, that what both enables us and forces us to share our reasons is, in a deep sense, our *social nature.* (p.135)

> [I]f our social nature is deep, in the sense that it is the nature of our reasons that they are public and shareable, then justifications of morality can and should appeal to it. So the kind of argument we need here is not one that shows us that our private reasons somehow commit us to public ones, but one that acknowledges that our reasons were never more than incidentally private in the first place. To act on a reason is already, essentially, to act on a consideration whose normative force may be shared with others. Once that is in place, it will be easy to show how we can get someone who acknowledges the value of his own humanity to see that he has moral obligations. (p.136)

One of these comments contains the germ of what I take the truth to be, namely 'To act on a reason is already, essentially, to act on a consideration whose normative force may be shared with others.' But we need to be more careful than she is in how to interpret this claim: there is an ambiguity in the notion of 'sharing' here. Think for a moment of instrumental rationality: let us say that John wants to become a millionaire. This gives him a reason to will the means to his end (for example, to work hard, or to marry an heiress). Now the normative force of the reasons that are *means* to his end must be shareable (it must be the case that, given his end, we can intersubjectively agree that *this* constitutes the best means to it), but it does not follow that the reasons in question are *agent-neutral.*

126

It is in the attempt to make this leap that Korsgaard introduces the idea of our 'deep social nature'. But the problem is that this seems to be in danger of going from an 'is' to an 'ought': the argument that Korsgaard constructs tries to go from something that doesn't have any normative significance (our social nature) to the publicity of reasons.[107]

4.5 *Conclusion*

After the collapse of Korsgaard's attempt to show how self-consciousness implies the validity of morality, we should attempt to salvage what is of value from it.

Charles Taylor presents one important way in which we might attempt to do this: he too invokes a transcendental account of human selfhood, but one which is closer to empirical psychological theories and generally less ambitious. He argues that for a life to have the sort of depth and value that is constitutive of a human life, an agent has to be hooked into a source of good outside himself, which is capable of providing the public reasons required.

I shall take up again, from my own perspective, the theme of our deep social nature in the discussion of recognition in chapter six and the theme of the public nature of moral reasons in the account of moral rules in chapter eight.

---

[107] In fact the figure whom she refers to, Wittgenstein, suggests that it is nothing to do with the fact that we are *social* animals that reasons are shareable: rather it has to do with the conditions under which normativity is possible. The key idea here is that private normativity is incoherent: if you take something to have normative force for you then this *ipso facto* commits you to the shareability of this normative force, on pain of your falling into mere wilful self assertion.

C h a p t e r   F i v e

Charles Taylor's Ethics of the Good

5.1 *Introduction*

At the close of the previous chapter, I argued that Korsgaard's attempt to reconstruct morality fails both because it fails to demonstrate that we *must* value our humanity, and because you cannot get from private to public reasons in the way she thinks we can. We can see Charles Taylor's moral philosophy as a way of remedying these deficiencies and thus giving a more convincing response to the Nietzschean critique of morality.

Charles Taylor's moral philosophy is immediately addressed to what he sees as the greatest challenge raised by Nietzsche: he agrees with Nietzsche that the question of the affirmation of life is of fundamental importance, and that, insofar as we cannot articulate why the life of a moral goodness is more worthwhile than any other life, morality is in trouble.

In arguing in this way, Taylor attempts to undermine methods of thinking about morality that focus, for example on autonomy, by insisting that the basis of moral thinking must be the good: that is, an image of a worthwhile life. On the basis of this understanding of the task of moral philosophy, he attempts to construct an argument that takes the following steps: i) we have a basic need to find our lives meaningful; ii) this need must be expressed in finding some things *good*; iii) where this sense of the good does not include the affirmation of the dignity of human beings, it is a result of a form of self-deception or unwillingness to see on the agent's part; iv) therefore if an agent sees things rightly, she will perceive the dignity of human beings and act accordingly.

I argue that his attempts to reorient moral philosophy, whilst convincing in part, go too far when they attempt to banish all other forms of normativity than that of the good: in attempting to remedy the one-sidedness of traditional moral philosophy, they go too far and we end up with a position that is just as lop-sided.

On the argument for morality, I show that we have no way of convincingly showing that the affirmation of the dignity of all human beings is in any way required: as we shall see, Taylor himself sees this to some extent, and suggests that a turn towards theism may be in order. I do not follow him in this, though I do pick up some of the themes that he wishes to push with his theism in the final chapter, though they are transplanted into a secular context.

The argument of this chapter takes the form of a reconstruction of and extrapolation from the whole range of Taylor's work. This is in part because Taylor philosophizes in a different vein from Korsgaard; he does not attempt to provide arguments that are knock-down in the sense that analytic philosophy (at least professes to) aspire to. Rather, his aim is to provide an account of human experience which we ourselves find compelling: Taylor favours this method over the more traditional knock-down approach, because believes that the attempt to provide foundational knock-down arguments is a *symptom* of the account of human beings that we need to go beyond if we are to explain how morality is possible.

5.2 *Situating Taylor*

For our purposes here we can vastly simplify Taylor's account as follows: up until the scientific revolution in the sixteenth century, the idea that there was an underlying order of things in the universe (in Taylor's phrase, an 'ontic logos') played a constitutive role in human life: it fixed the guiding principles of science and of ethics and of the nature of man. Moreover, it gave a unity to the overall outlook: there was, for example, no fact-value

dichotomy. Science and axiology were closely intertwined, for the same underlying logos served both to explain how things were and how they ought to be. Knowledge too was understood in terms of this world view: rather than as a *representation* of how things really were, it was understood *participationally,* 'being informed by the same *eidos,* the mind participated in the being of the known object, rather than simply depicting it.' (1987, p.467)

The coming of the scientific revolution destroyed Aristotelian science, and with it, swept away the idea of an ontic logos, which was replaced by the conception of the world as inert, evaluatively neutral. Given that the ontic logos had played at constitutive role in the notions of science, ethics and of man, these all had to undergo extensive reformulation.

The key switch here, Taylor suggests, is in the underlying conception of reason: '[R]ationality is no longer defined substantively, in terms of orders of being, but rather procedurally, in terms of the standards by which we construct orders in science and life.' (1989, p.156) The flip side of this new idea of procedural reason is what I would call an *ethic of disengagement,* which holds that we ought, so far as is possible, abstract from our human situatedness in the world, so that we can discover the truth. I speculate that the ethic of disengagement comes out of the idea of procedural reason by something like the following chain of thought:

> The only thing that is trustworthy is reason. But reason is *formal,* in that it depends on following *procedures* which are constitutive of it. Hence the things over which reason ranges, and which it can pronounce rational are not intrinsically trustworthy: we must be suspicious of them. So it is best to disengage ourselves from them as far as possible. [My reconstruction]

For Taylor, the two most important ramifications of the switch from substantive to procedural conceptions of reason relate to its conceptions of knowledge and of ethics.

Knowledge comes to seem both philosophically more important, and much harder to attain than it had previously. The general model that procedural reason uses for its account of knowledge is that of an inner representation of an outer reality: what makes a particular representation *knowledge* as opposed to a belief that just happens to coincide with the truth is that the belief has been derived by the application of a *reliable method*. Now, this new world view ushers in a threat that was not present before: once we consider what is in our minds to be merely a *representation* of reality, then the problem arises of how we can be sure that our representations match up with the way things really are. Scepticism becomes a constant threat – not, of course, because anybody actually believes it – but because this representationalist construal of knowledge makes it difficult to see how our 'knowledge' could be adequately grounded.

In the ethical sphere, the switch to procedural reason creates an inversion: where the focus in ethics had been the good (what is most worth aiming at in action and in human life), it becomes the right (the standards that are appropriate to the regulation of human action). There are two different specifications of ethics compatible with this conception of procedural reason, depending on the degree of power which we accord to reason. If the power of reason is thought to be slight, then we have something like the rational economic man; if it is thought to be great, we have a Kantian model.

5.3 *Taylor's critique of the disengaged worldview*

Much of Taylor's philosophical effort goes into trying to undermine this 'disengaged' way of looking at things; his clear assumption being that it is this ethic of disengagement that has created most of the problems and that overcoming it will *ipso facto* put us a good distance towards finding a solution.

Taylor takes it that the conception of knowledge is the key element that must be overturned in a critique of this worldview. The obvious reason for this is that, according to Taylor's philosophical history, it is its conception of knowledge that gives this worldview its particular structure. More subtly, Taylor's philosophical history has an ambiguous relationship to Hegel, which is perhaps shining through here.[108] Hegel argues that what is counted as knowledge always plays this structural role in formations of consciousness and that it is through perceived inadequacies in the structures which had, up till then, counted as knowledge that cognitive improvement occurs.[109] Certainly, Taylor intends his account knowledge to play a constitutive role in his ethical thinking.

5.3.1 *The argument from agent's knowledge*

The ideal of disengaged knowledge is to be undermined by an appeal 'agent's knowledge'. 'Agent's knowledge' is here to be understood as a grasp that we have, in virtue of being agents, of our own agency: 'As those effectively engaged in the activities of getting to perceive and know the world, we are capable of identifying certain conditions without which our activity would fall apart into incoherence.' (Taylor 1987, p.475)

The disengaged world view, of course, sees the idea of a *representation* as basic in knowledge. Whether linguistically, pictorially, or in some other way, there is a representation that things are thus-and-so, (for example, *that the cat is on the mat*) which may or may not match up with the way things really are.

What Taylor's critique aims to show here is that the idea of a representation is *not* basic, in that it *rests on* a hitherto unacknowledged engagement with the world; showing this through reflection on the transcendental conditions of our having the sort of agent's knowledge that we do in fact have. In

---

[108] When Taylor grapples with the question about his methodology in chapter 12 of *Sources of the self* he (in my opinion) fails to recognize how Hegelian it is.

[109] *Phenomenology of Spirit,* Introduction.

Taylor's words, that 'What you get underlying our representations of the world – the kind of things we formulate, for instance, in declarative sentences – is not further sentences – is not further representations but a certain grasp of the world that we have as agents in it.' (1987, p 477)

Taylor takes the adequate exposition of this thesis to be the key advance in twentieth century philosophy: in this he thinks that Heidegger, Merleau-Ponty and Wittgenstein[110] have done the key work, and it remains for us only to defend their insights against misunderstandings engendered by the ideal of disengagement. Where the 'epistemological construal' takes the idea of a representation as basic and takes the fundamental question to be how we can know that our inner representations match up with outer reality, Heidegger shows that 'the condition of our forming disengaged representations of reality is that we be already engaged in coping with our world, dealing with things in it, at grips with them.' (1987, p.476)

This realisation is supposed to undermine the whole world view and its accompanying ethic of disengagement. It turns out that being at grips with the world is the precondition for even being able to have representations that we are capable of disengaging from, so the totally disengaged agent does not make sense even as an ideal. Rather we must accept ourselves as situated, social beings, and realise that it is in virtue of the situatedness that we are able to think at all, as Taylor puts it, 'We cannot turn the background from which we think into an object for us. The task of reason has to be conceived quite differently: as that of articulating this background, "disclosing" what it involves. This may open the way to detaching ourselves from or altering part of what has constituted it – may, indeed, make such alteration irresistible; but only through our unquestioning reliance on the rest.' (1987, pp. 477-8)

Remarks on this argument

---

[110] See *Being and Time, The Phenomenology of Perception* and *Philosophical Investigations* respectively.

1: Is the argument question begging?

The difficulty with this argument is that it attempts to, so to speak, pull itself up by its own bootstraps. It runs a transcendental argument from a certain conception of what is open to us via our agent's knowledge, and hence installs these things as basic. But in doing so, it is not clear that it has *refuted* the representationalist picture. Rather, it might just be arguing from different premisses. What assurances can it give us that it is not just begging the question against the representationalist construal?

Taylor has the beginnings of an answer to this question, but it seems to me inadequate: this answer is expressed in embryo in the following explanation

> By "transcendental arguments" I mean arguments that start from some putatively undeniable facet of our experience in order to conclude that this experience must have certain features or be of a certain type, for otherwise this undeniable facet could not be. (1972, p.151)

The reason, then that the argument from agent's knowledge is not question-begging, according to Taylor, then, is that it starts from facets of experience that are undeniable.

The important question is how we are supposed to know these undeniable elements of experience: is this knowledge supposed to be immediate, or is it supposed to be mediated through philosophical theories? If it is supposed to be known immediately, then we need some argument as to how the structure of experience can be known in an unmediated way: if its mediated, then it becomes dubious how *basic* such knowledge could be said to be.

2: The Heideggerean conclusion about reason seems to rely on a confusion between the nature of the knowing agent and the methods/procedures appropriate to reason.

We might agree that each and every human being is situated within a community, and that it is only in virtue of being so situated that he or she is able to reflect. But it does not follow from this that the standards appropriate to reason are *constituted* by our engagement with the world. I shall argue below that for reason to be possible at all, it cannot be so constituted by our engagement with the world. But for the time being we should rest content with pointing out that there is nothing to stop the representationalist re-appropriating what Taylor has said; admitting our situatedness, but still arguing there are compelling reasons for us to strive for knowledge that is as far as possible *absolute.*

If the account of knowledge was to provide a foundation for what he wanted to say in ethics, then it has been unsuccessful; it should at least have served the lesser function of filling out the background necessary for an understanding of Taylor's work.

5.3.2 *The argument against a purely procedural ethics*

Taylor also has another argument, wholly his own, which we can think of as parallel to the critique of the disengaged subject of knowledge. This one attempts to attack the idea of a disengaged subject of practical reason. Taylor equates the attempt to construct an ethics out of the picture of a disengaged subject of practical reason with purely procedural ethics. A purely procedural ethics would be so constituted that:

1) Morality is exhausted by the notion of right action

2) Right action is exhausted by following procedures that are in themselves rational.

Taylor's aim is to show that such an ethics would be incoherent, and to the extent that real ethical positions (e.g. utilitarianism, Kant's ethics) actually resemble it, they are either incoherent, or trade on values that they cannot, by their own lights admit to.

The crushing problem which is supposed to beset any purely procedural ethics is that, within its own terms, it cannot give an account of the *value* of rationality, of the superiority of living as the rational person does.

To make this point fully clear, we must make a distinction: a purely procedural ethics answers questions about *individual* actions, by referring us to the concept of rationality (for example, "Why must I study?" – "Because, given that you will the end [passing the exam], it is rationally required that you pursue the means [studying]"). But no such answer is available if we ask the question about rationality itself – for in this case to refer us back to the requirements of rationality is question begging and circular (for example, "Why should I do what is rational" – "Because that's what rationality requires")

The ethics of autonomy threatens to force an inarticulacy on us about the ultimate source of the reasons for our actions: we seem to be left with no reason to think that being rational *matters* – that is, we seem to leave the door open for someone to merely reply 'so what' when we admonish them that they are being irrational. What is worse, we would be left with no account to give *ourselves* of why rationality mattered. In essence, the problem is this: a disengaged, purely procedural ethics only gives us an abstract structure of rules and requirements and cannot in itself demonstrate the appropriateness of these rules and requirements for the regulation of the actions of embedded human beings like ourselves.

Taylor takes it that these reflections show that an account of the *good* is required at the base of ethics. Taylor uses 'the good' as a term of art, to describe a vision of a worthwhile, *higher* form of life that the agent takes to be normative and which is motivationally efficacious for him.[111]

The reason why only the good could be suitable basis for ethics is that it is only with the good that we reach a justification that a human being living

---

[111] I discuss Taylor's account of our psychological economy and its relation to the good later in the section on Taylor's philosophical anthropology.

in the world will find satisfying. Hence, we should conclude that all ethical discussion must be ultimately answerable to an account of the good: as Taylor puts it, 'The chain of explanations must be anchored somewhere in our intuitive grasp of what is at stake.' (1985b, p.62)

Taylor suggests that this sound point has overlooked, because ethics that have presented themselves as procedural have in fact relied on a conception of the good, namely the conception of human beings as possessing a *dignity* in virtue of their rational nature. The good was not banished by procedural ethics, but *displaced inwards* into the idea of the dignity of human beings:

> If rational control is a matter of mind dominating a disenchanted world of matter, then the sense of the superiority of the good life, and the aspiration to attain it, must come from the agent's sense of his own dignity as a rational being. (1989, p.152)

However, procedural ethics have either not admitted to this conception of the good (as in utilitarianism), or have admitted it, but urged that it is the subjective effect of the moral law (Kant).

Any proposed purely procedural ethics is left with a dilemma: either it comes clean, and admits that it is powered by an account of the good, namely that human beings owe their dignity and importance to their power of reason. But this would involve it in a contradiction: it posits autonomy and disengaged reason as its first principles, but if the status of these *as first principles* is contingent on the subscribing to a particular vision of the good, then it turns out that autonomy is grounded in heteronomy, and the ideal of disengaged reason in an engagement with the world.

Or, if it attempts to eschew its account of the good, then, as we have seen, its requirements remain a mere abstract structure without any necessary applicability to human life.[112]

---

[112] For this argument to go through against Kant's position, Taylor would require some way of demonstrating that the feeling of respect is *not* just the subjective effect of the moral law on a finite

Remarks on this argument

I think that this argument tells us something useful and important, about what is left out by some forms of utilitarianism and by Kant's official account of his own views. And I think that we should agree that when it comes to morality, we must take the good to be in a certain sense prior to the right.

However, Taylor overreaches in his estimation of what is achieved by this argument: he takes it to imply that ethical theorizing should concern itself with the good rather than the right. If this is followed through, then right is robbed of the intrinsic normativity that it purports to have; rather, moonlike, it borrows its light from the sun of the good.

However, this stronger conclusion does not follow. For it does not seem to be generally true that all standards of practical rationality must be anchored in a conception of the good. Take means-end reasoning, for example. Dreier (1997, p. 93) considers the following case:

> We tell [Ann] that she ought to take a prep course for the Law School Admissions Test. She asks why. We point out that she wants to raise her chances of getting into a competitive law school, and she can raise her chances by taking the prep course. She admits as much, but still isn't motivated to take the prep course. So we cite a rule, the means/end rule:
>
> M/E: If you desire to ψ, and you believe that by φ-ing you will ψ, then you ought to φ.
>
> Now suppose that Ann agrees that this rule does indeed instruct her to take the prep course, given what she believes and desires, but she shrugs at the rule. She doesn't accept it.

Taylor's position would commit him to arguing that what lies behind the validity of the means-end rule is a vision of the good, namely of a strong-

---

will, but is in fact a vision of the good that plays a foundational role in Kant's theory. If he wished to make this argument, he could avail himself of the arguments we put forward in chapter two concerning the relationship between universal law and the humanity formulations of the categorical imperative.

willed, effective person who takes the necessary steps to secure what she desires. But there seems to be no good reason to follow him in this: the means-end rule is more naturally understood as having a *constitutive* status: on this reading, the attempt to ask *why* we should follow the means-end principle just shows a basic lack of understanding of practical reason, for without it, one has no grip on the idea of a reason for action at all.[113]

In the case of moral rules, we can agree with Taylor that they must be put forward within a wider context of an account of the good: but it does not follow that the normativity of moral rules is reducible to that of the good. For it is quite possible to have a position (such as the one I shall be arguing for in the final chapter), in which the good provides the context in which the attempt to construct moral rules that all can agree to makes sense, but where the validity of moral rules is to be decided by a dedicated and separate procedure.

Taylor's error here is parallel to that in the case of the account of knowledge: just as engagement need not exhaust theoretical reason, so it need not exhaust practical reason. Indeed, there are reasons that spring from within the nature of reason itself that should lead us to be wary of attempting to tie the correct exercise of reason too closely to the boundaries of our all-too human engagement. While reasoners are engaged agents operating within social horizons, reason *qua* reason seeks a validity beyond all such horizons. As Putnam explains it, 'Reason is, in this sense, both immanent (not to be found outside of concrete language games and

---

[113] Dreier puts the point as follows: 'What would count as a reason, by her lights? As long as she accepts M/E, we know what would count as a reason: some belief that by following the rule to be justified she would achieve some end she desires. ... If you can't draw the practical inference, not even the fundamental M/E kind, then nothing counts as a reason for you. This is why M/E has a kind of ground-level normative status. I think it counts as a categorical imperative, too. Of course, the particular reasons that M/E generates are all hypothetical reasons. But M/E is not hypothetical. Its demands must be met by you, insofar as you are rational, no matter what desires you happen to have.' (Dreier 1997, pp. 95-6). For other useful recent writing on instrumental reasoning, see Korsgaard (1997), Hampton (1998). For further thoughts on constitutive arguments about practical reasoning, see Velleman (1992), (1996), (1997) and Railton (1997).

institutions) and transcendent (a regulative idea that we use to criticise the conduct of all activities and institutions)'. (Putnam 1982, p. 228)

In so far as an ethical norm or a methodology for thinking purports to be rational, it must be able to survive critique *by reason*. But reason itself is also subject to critique by reason. In so far as anything fails such a test, it only approximated to rationality; or perhaps was a downright mistake. We cannot state in advance that the results of such testing will be the exclusive upholding of the picture derived from our basic engagement with the world, 'for it is indeed absurd to look to reason for enlightenment, and yet prescribe beforehand which side she must necessarily favour'. (Kant *C1,* A747/B775)

5.3.3 *The argument from the phenomenology of experience*

We have so far seen no reason to follow Taylor in believing that the normativity of the right is borrowed from that of the good. But he also has another argument to this conclusion, based on the phenomenology of experience.

The argument turns on making us look and see the way in which concepts like 'shame' operate. We are supposed to appreciate i) that such concepts presuppose an engagement with the world; ii) the feelings correlative to such concepts are expressions of some underlying sense of what is important[114] (i.e. a notion of the good); iii) (generalizing this conclusion) it is only through feeling that we can gain access to what is important (to the good); and iv) morality too must be explained in these terms.

i) Taylor argues that engagedness is necessary for shame in the following way:

---

[114] Taylor uses 'what I feel as important' in preference to discussion of 'the good' in 'Self-interpreting Animals' (1985b), from which I draw this argument. This does not represent a difference in position, but only in focus. In 'Self-interpreting Animals', he is considering solely what having a notion of the good looks like from the inside, and finds the notion of importance helpful.

> Shame is an emotion that a subject experiences in relation to a dimension of his existence as a subject. What we can be ashamed of are properties which are essentially properties of a subject. This may not be immediately evident, because I may be ashamed of my shrill voice, or my effeminate hands. But of course it only makes sense to see these as objects of shame if they have for me or my culture an expressive dimension: a shrill voice is (to me, to my culture) something unmanly, betokens hysteria, not something solid, strong, macho, self-contained. It does not radiate a sense of strength, capacity, superiority. Effeminate hands are – effeminate. Both voice and hands clash with what I aspire to be, feel that my dignity demands that I be, as a person, a presence among others. (1985b, p.53)

Shame involves us in engagement with the world in two ways: first, it is an emotion that can only be had by an engaged agent, as it is experienced relative to the agent's own understanding of himself as an agent (or, in Taylor's terms, it is 'subject-referring'). Second, if we are to understand or justify the actions that shame motivates, this too can only be done from the perspective of an engaged agent: this is true both on a conceptual level i.e., that if we wish either to explain actions done from shame or to explicate them we have to refer to things like 'our sense of dignity, of worth, of how we are seen by others – which are essentially bound up with the life of a subject of experience.' (1985b, p.54) It is also true on the level of producing actual interpretations: to fully understand an action done from shame we must have a sense of the form of life which that feeling of shame articulates.

ii) the feelings correlative to such concepts are expressions of some underlying sense of what is important (i.e. a notion of the good).

I abstract this premiss from what Taylor says about 'subject-referring' properties.

iii) (generalizing this conclusion) it is only through feeling that we can gain access to what is important (to the good)

141

Taylor formulates his argument for this claim in response to an imagined criticism that '[W]e also see our selves as being aware of what is important to us as humans in other ways... We should distinguish what we feel as important or valuable, we are often led to say, from what we know rationally to be' (1985b, p.60) Taylor's response is that this rests on a misleading characterization; for what I 'know rationally' to be valuable is *also* based in feeling:

> If I want to say that I know certain things to be truly important: one should be generous even to blackguards; or that the only thing one should really be ashamed of is being untrue to oneself; or that acting out of spite is always bad – even though my gorge rises at the thought of helping that cad, I feel ashamed at not making the football team, and I cannot resist lashing out at my successful rival – I am not just opposing feeling to reason. For I would not 'know' that one should be generous, and so on, unless I was moved in some way: perhaps I feel remorse when I have delivered myself of a spiteful attack; or feel self-contempt at my lack of autonomy when I allow myself to feel shame at not making the football team; or feel morally inspired by the ideal of universal generosity. If I were quite impervious to any such feelings, these norms and ideals would carry no weight with me; I would not even be tempted to subscribe to them, and would not describe myself as 'knowing' that they were true/valid. (1985b, pp. 60-1)

The key idea here is that valuing *x* has an internal relation to being moved by it; hence to talk of 'knowing' that something is valuable or important is misleading, if it is taken to imply that this knowledge is unrelated to being moved. '[W]e can have no dispassionate awareness of the human good.' (1985b, p.62)

There are two problems with this stage of the argument: the first is that the proposed necessary link between motivation and the human good is suspect. Taylor allays this difficulty somewhat in the later *Sources of the Self,* where he spells out this relationship in a more nuanced way:

> Our acceptance of any [good] is connected in a complex way with our being *moved* by it. ... We may accept something as a good although we are relatively unmoved by it, because at the lowest, we think very little about it and glide along in conformity with our milieu; or because we revere and look up to established authority; or perhaps best, because we choose certain figures as authoritative for us, sensing in them that they are moved by something authentic and great, even though we don't fully understand it or feel it ourselves. But through all these complex chains of intermediation, the connection between seeing the good and being moved by it cannot be broken. (1989, pp. 73-4)

Second, it remains unclear how 'the human good', defined in terms of the individual's feeling that something is *important,* is related to what philosophers would usually describe as the good for a human being. 'The good for a human being' would include, no doubt, some notion of what is *important,* but it would also involve more mundane elements, such as the basic needs of food, shelter, warmth, association and so on, which while we have them, seem relatively unimportant. To take 'the human good' in Taylor's sense to be everything in ethical inquiry seems to risk romanticising our predicament in such a way that ethical inquiry is cut off from the needs deriving from the sort of biological beings we are.

In fact, we can quite easily think of cases in which a form of life associated with the reverence of a particular good is actually harmful from a biological point of view (primitive initiation ceremonies spring to mind). It is ludicrous to treat our feelings of what is important as enjoying a complete autonomy from the actual conditions of human life: I do not see how anyone could fail to see how the fact that holding certain values is physically bad for us constitutes an objection to them.

This objection is sufficiently obvious that one would have thought that Taylor would have some answer to it; however, I have been unable to find a very clear one in the corpus of his writings. What I suspect he would say, if pushed, is that the 'no dispassionate awareness of the human good' thesis

does not commit us to merely contemplating our own feelings; other sorts of knowledge and expertise can of course be brought to bear on the subject of the good. He would stress, however, that it is in our feelings of what is important that the ultimate court of appeal lies.

iv) We have already criticised Taylor's conclusion that because the right presupposes an account of the good, it therefore derives its normativity from the good. Taylor attempts to give another argument for his conclusion, based on the reading of the phenomenology of moral obligation.

Taylor bases his claim on an explanation of the experience of moral obligation. He suggests that we take, as a paradigm case, an experience something like the one that the good Samaritan had, that of 'coming across someone in trouble and feeling called upon to help'. (1985b, p. 57) His reading is that this experience involves a feeling of what is of importance, which is related to our conception of ourselves:

> For I do not just feel desire to help this man. Indeed, I might feel no such desire in the usual sense of the term. But I feel called upon to help him. And I feel called upon *qua* rational being, or moral being, or creature made by God in his image, in other words capable of responding to this like God, that is, out of agape. The obligation does not lie on an animal nor, in another way, on an idiot, nor an infant. (1985b, p. 58)

Here we might agree with Taylor's reading of what it is like to feel a moral demand, but without the argument which claims that the normativity of the right is derived from that of the good, nothing follows about validity of moral demands – certainly it does not follow that the validity of a moral demand is to be judged by its ability to move us.

5.3.4 *Principle of the best account*

Last of all, Taylor attempts to ground his ethical philosophy with what he calls the 'principle of the best account'. Taylor lays out his argument in the form of rhetorical questions:

> What better measure of reality do we have in human affairs than those terms which on critical reflection and after correction of the errors we can detect make the best sense of our lives? 'Making best sense' here includes not only offering the best, most realistic orientation about the good but also allowing us best to understand and make sense of the actions and feelings of ourselves and others. (1989, p.57)

> What we need to *explain* is people living their lives; the terms in which they cannot avoid living them cannot be removed from the explanandum, unless we can propose other terms in which they could live them more clairvoyantly. We cannot just leap outside of these terms altogether, on the grounds that their logic doesn't fit some model of "science" and that we know a priori that human beings must be explicable in this "science". This begs the question. How can we ever know that humans can be explained by any scientific theory *until* we actually explain how they live their lives in its terms?

> ... The terms we select have to make sense across the whole range of both explanatory and life uses. The terms indispensable for the latter are part of the story that makes best sense of us, unless and until we can replace them with more clairvoyant substitutes. The result of this search for clairvoyance yields the best account we can give at any given time, and no epistemological or metaphysical considerations of a more general kind about science or nature can justify setting this aside. The best account in the above sense is trumps. (1989, p.58)

More rigorously, we could rewrite the argument as follows:

We take as a premiss the claim that the function of our evaluative terms and practices is to help us to live wisely and insightfully as human beings. Anything that is designed to fulfil a function F enjoys a certain autonomy,

in the following sense: *qua designed to F,* it is criticizable only on the grounds that it does not F well. Thus, a rain water butt qua rain water butt is criticizable only on the basis of its failure to catch and store rain water. If we extend the same principle to our evaluative terms and practices, they are criticizable as such, in part or in whole, only on the basis that they fail to help us to choose wisely and insightfully as human beings. For example, we would have *lost*, ethically speaking, if in attempting to make our evaluative practices more 'rigorous' or 'scientifically acceptable', we thereby reduced our ability to judge wisely and insightfully.

It follows that something could only count as an improvement of our evaluative practices if it enables us to do better what our evaluative practices attempt to do. Our evaluative practices are therefore *autonomous,* in the sense that no change can be imposed on them from the outside; rather it must meet with the approval of these practices themselves.

This should give us important additional reasons for being suspicious of any attempt to reduce our discussion of ethical matters to facts that can be known independent of specifically human sensibilities. A scientific theory's excellence *as science* does not give any intrinsic reason why it should increase ethical insight; rather before we rectified our evaluative practices on the basis of a scientific theory, we would have to check that the scientific theory did in fact better fulfil the function of our evaluative practices. In fact, as we have seen above nothing which is put in absolute terms can tell us anything about the human good; what we are seeing here is that it is dubious to imagine that science could even put background constraints on acceptability in ethical thinking. Hence the commonly held background assumption that science is a higher or more rigorous form of knowledge than 'mere' common sense must be abandoned in the case of ethical thinking.

The corollary that we should draw is that in discussing the human good, to be truly philosophical – that is, to adopt the methodology appropriate to the task in hand – is to attempt to make sense of our lives as we live them rather than to adopt a model of rigour borrowed from outside discourse concerning the human good.

Criticism of this argument

What Taylor is gunning against with the principle of the best account is reductive movements like behaviourism, which, on the basis of methodological assumptions attempt to bypass the inner logic of our deliberative practices.[115] But he oversteps the mark: the principle of the best account amounts to an attempt to hive off our deliberative practices as something *quite separate* from what can be discovered by science. This I would submit, is both an over-reaction, and a severe methodological mistake. It is an over-reaction because behaviourism was *bad science:* it was quite incapable of explaining human behaviour because the theory contained insufficient resources to explain how an animal can act so as to mould its environment in the absence of a stimulus.

It is a severe methodological mistake because in its attempt to carve out a specific niche for ethical thinking, it forgets a far more fundamental duty of philosophy, namely to mediate *between* different systems of thought which claim a genuine cognitive grasp on the way things are, but which produce results that are either incommensurable with one another, or are at odds with one another.

If we consider the law of non-contradiction, then it should be apparent why philosophy should have to fulfil this task: the law of non-contradiction states that the same attribute cannot at the same time belong to the same subject in the same respect.[116] Now, presumably in claiming knowledge of something, we are claiming to cognize it as it really is. Now, it is clear that

---

[115] See Taylor's first book (1964), which is a critique of behaviourism.
[116] Aristotle, *Metaphysics* IV.3; 1005^{b}19

what one system claims as knowledge cannot simultaneously be claimed by a different system to be *not* knowledge in the same respect, but yet both be correct. Hence, in all our dealings with science and with ethical theorizing, we should remember that all the truths there are must cohere with one another. Even though our knowledge may be highly fragmented, we should never lose sight of the fact that it is incoherent to suppose that the pieces in front of us are not all part of one jigsaw.

This is, of course, not to say that it is a requirement of reason that all truths can be reduced to some sort of common currency; instead it is to affirm that reason requires us to give some account of logical space which explains how everything that we take to be true can indeed be true simultaneously – which will, for example, allow us to distinguish different explanatory levels. In so far as the principle of the best account fails to take this into account it is founded on bad philosophy.

We have seen that the grounds of which Taylor attempts to put forward his approach to ethical theorizing as the only possible one are flawed. It remains to see how Taylor applies this conception of ethical theorizing.

5.4 *Taylor's philosophical anthropology*

Taylor's philosophical anthropology attempts to lay the foundations for morality, by arguing that our best account of human beings will show us that the centre of each person's identity is an account of the good which commits them to public reasons. It attempts to tighten up all the loose formulations we have made so far in terms of the good, and 'importance' into a unified theory.

Taylor starts from what he takes to be the common sense idea that there are certain things that it is good to be or good to love. Thus, we might say it is good to be generous or trusting, and good to love the truth. Actions obviously will follow from such an adherence; but when we admire

someone, for example, who is generous even to his enemies, we are evaluating the *person* or *the quality of will,* that lies behind these actions.

Taylor first attempts to give a more perspicuous explanation of these 'things that are good to be or to love', and second to show, via a transcendental argument, that having things you believe it to be good to be or to love is a necessary condition for undamaged human personhood.

5.4.1 *The descriptive part*

Taylor invents the term 'strong evaluation' to attempt to explicate what lies behind the claim that there are certain things that it is good to be or to love. He has made several attempts to elucidate his conception of strong evaluation, and it has been argued that these attempts are not all mutually consistent.[117] Here I quote the attempt which seems to me to be most helpful:

> [An] end is strongly valued when its being an end for us is not just contingent on our happening to desire it or need it, when in other words, we allow that we would be lesser beings if we should cease to want or need it. The ice-cream cone I now desire is weakly valued, because should I lose my interest in it, it would no longer have any claim on me. But I do not think of my commitment to Amnesty International in these terms. Someone who refused to contribute to this cause because they were not 'into' torture victims this week would be thought unbearably frivolous. Their change of interest does not reduce the claim, but rather shows them in a poor light. (Taylor 1995, p. 134)

This is a little vague and impressionistic, but we can tighten it up into the following four theses: An end is strongly valued when it has the following properties:

---

[117] See Taylor (1977), (1985b), (1985c), (1989) and (1995). This complaint of inconsistency is made by Owen Flanagan (1990). Flanagan does not refer to the account I shall examine, (Taylor 1995), which was published after his article.

i) The end is a standard against which one can be judged regardless of the desires I now have.

ii) The end is such that it draws people towards it because they see it to be good.

iii) Being motivated by this end is being motivated by the good, and is 'higher' or praiseworthy.

iv) Correlatively, failing to be moved by what is strongly valued can be an adequate reason for contempt.

As is implicit in the above formulation, Taylor believes strong evaluation to be intrinsically *contrastive* in nature. He does not explain fully why this is so; however, Nicholas Smith, in his perceptive and sympathetic account of Taylor, gives a good explanation:

> Mattering, Taylor informs us, is only intelligible as a background of qualitative distinction; if everything mattered the same, if anything mattered, nothing would. What matters *makes a difference*, [hence] its articulation requires qualitative distinctions between the worthwhile and the worthless, the significant and the trivial, the fulfilling and the vacuous. (Smith 1997, p. 37: my italics).

Once we switch to the terminology of 'mattering', rather than 'what it is good to be', then everything becomes clearer: we are able to see that the notion of *mattering* already entails that of a 'higher' that which matters as opposed to a 'lower' – that which does not matter.

Mattering must be considered holistically because mattering is a comparative concept (for example, like tall) rather than an absolute one (for example, having a height of six feet). (The distinction being that the height of other men makes a difference to whether this man is tall or not, but it makes no difference to whether he is six feet tall). Just as we can only say that a man is tall against some background understanding of the normal height for a man, so we can only say that something matters against a

background understanding of what is important. It is this general understanding of what is important that is a 'notion of the good'.

That a person is a strong evaluator ties her into a community of valuers, Taylor argues. His reasons for this are threefold: first, that one can only grow into an adult human being in the context of a community – by taking up the values of that community and internalizing them. Second, once you are a fully functional adult, you can only progress by working on the framework you find yourself in; repairing the ship whilst it is at sea. Third, you can only make progress on improving this framework through dialogue; whether with others who are physically present; positions one encounters in books or through hearsay or through inner dialogue in which you yourself play the parts of the people you are arguing with. Each way, the progress you are making is within a community of valuers; you are presenting your notion of the good to a community for its approval (even if it is only the notional one of an inner dialogue).

Note that what we have said about strong evaluation so far is compatible with two distinct readings of the ontology it calls into play. We could call these the *subjectivist* and the *objectivist* readings. Obviously the *epistemology* of strong evaluation is subjectivist, as we can only be strong evaluators if we have the correct specifically human sensibilities and grown up in a community. But this leaves open the question of whether there is something that *lies behind* the strong evaluations we make. There are two possibilities here: the subjectivist one would state that nothing lies behind our strong evaluations. Our strong evaluations articulate the form of life of a community; they are part of a way of living in the world, and *that is all they are*. The objectivist reading would agree that our strong evaluations articulate the form of life of a community, and that they are part of a way of living in the world, but it would say that there is also something that lies *behind* any set of strong evaluations – an objective

good – that any particular set of strong evaluations is an inadequate attempt to articulate.

Reading between the lines, it is quite easy to see that Taylor believes in the objectivist reading. But, judging from his way of proceeding, he is also aware that it would not be a valid philosophical move to merely *assert* that there is an objective good resting behind our strong evaluations. Rather he must provide an *argument.*

This leaves him, it seems, with two possible strategies: i) to argue for morality whilst only presupposing a *subjectivist* account of strong evaluation, or ii) to give grounds, from within the nature of the strong evaluations people actually have, for thinking that the objectivist account of strong evaluation better characterizes the way things are.

I'm not sure that Taylor has clearly distinguished strategy i) from strategy ii) in his mind. Anyway, I shall show below that i) is fraught with problems, and that, given Taylor's conception of normativity, strategy ii) requires the success of strategy i), and hence is unworkable.

5.4.2 *Strong evaluation and human life*

Before this, Taylor must show that strong evaluation can be taken to be universal for human beings. Both strategy i) and strategy ii) require this: i) presupposes that each person has some strong evaluations that we work from to get to morality; ii) requires it because it needs an inseverable link between objective value and motivation.

Taylor's approach is to try to show that strong evaluation is a *transcendental condition* of human agency; as he puts it,

> Living within ... strongly qualified horizons is constitutive of human agency, that stepping outside these limits would be tantamount to stepping outside what we would recognize as integral, that is, undamaged human personhood. (1989, p. 27)

Smith provides a reconstruction of this argument that is more perspicuous than Taylor's original, which I will for the most part follow. Smith suggests that we can best understand the argument with reference to a distinction that Paul Ricoeur makes between two different senses of identity:

> On the one hand, one's identity is that which makes one the same, it signifies permanence in time, and it has that which differs, the changing and the variable, as contraries. This sense of identity corresponds to the Latin term *idem.* On the other hand, there is the sense of identity that corresponds to the Latin *ipse,* one that is preserved in the English word 'ipseity'. Synonyms for ipse-identity include individuality and selfhood. Ipseity signifies *my*self, my selfhood in contrast to sameness. (Smith 1997, p. 50: the distinction is made in Ricoeur 1992)

Now, Smith suggests, we should take Taylor's argument to be a transcendental one, about the possibility of *ipse* identity.[118] This illuminates Taylor's argument considerably. Taylor defines identity as follows

> My identity is defined by the commitments and identifications which provide the frame or horizon within which I can try to determine from case to case what is good, or valuable, or what ought to be done, or what I endorse or oppose. In other words the horizon within which I am capable of taking a stand. (1989, p. 27)

This is the role that an ipse identity must fulfil: but this is already to presuppose strong evaluation, for our identity 'only plays the role of orienting us, of providing the frame within which things have meaning for us, by virtue of the qualitative distinctions it incorporates.' (1989, p. 30)

If this were supposed to be an argument in the traditional sense then it would be quite obviously question begging; but we should be more charitable to Taylor and interpret it in the light of his attempt to provide the best account of human life. When we think of it in this way, there are

---

[118] Hence in terms of Taylor's definition of a transcendental argument that I quoted earlier, this transcendental argument takes our ipseity to be a 'putatively undeniable facet of our experience'.

additional reasons that can be adduced in support of the connection between ipseity and strong evaluation.

First, Taylor urges, is difficult to imagine how we could make any sense of the idea of an ipseity *which did not involve* strong evaluation. The problem here would be that 'our identity is what allows us to define what is important to us and what is not. It is what makes possible these discriminations, including strong evaluations. It hence couldn't be entirely without such evaluations.' (Taylor 1989, p. 30)

Second, the phenomenon of an identity crisis can be used to give us insight into the conditions of ipseity. If we take it that 'To suffer an identity crisis is to be incapable of telling why a life should be led one way rather than another,' (Smith 1997, p. 39) then identity crises will prove to be enlightening for our purposes, because this sense of not being able to discriminate between what is more and what less worthwhile is experienced *by the agent* as 'an acute form of disorientation'. (Taylor 1989, p. 27) But it is only possible for the agent to experience the identity crisis *as* an extreme form of disorientation on the assumption that 'the question of what is really of more or less importance, worthwhile or fulfilling *demands* an answer even if we are not in a position to give one.' (Smith 1997, p. 39)

So, either a human being has an orientation to the good which gives them a place to stand on what they find important; or they lack one, and so suffer a feeling of disorientation. Hence we are entitled to conclude that human identity involves an orientation to the good and hence strong evaluation.

Remarks on this argument

I like the conception of the self that comes out of this transcendental argument, as defined by that which it takes to be good. I would be happy to take it a moral or rational ideal. But I am sceptical of the claim that human selves must be so constructed. Taylor's account implies that one leads an

authentic life through following the strong evaluations that one adheres to. I would agree that this is the ideal, but I do not agree that everyone has the necessary self-clarity or articulacy to do so. It seems far more plausible to suggest that most of us have only made small steps towards this goal.

While we are still distant from this goal, there can be a massive chasm between the values that one would need to assert to live at anywhere near one's optimal capacity, and the ones that one currently takes to be authoritative for oneself. Think, for example, of someone who is gay but has grown up in an environment in which homosexuality is seen as evil and unnatural. Now, suppose he has internalized these values (that is, imported them into his self as a *structural* element). He now has a problem: he is drawn to what he would simultaneously revile as evil. There are various ways the situation could develop from here. I will very briefly consider some of them.

1. Authenticity. By articulating for himself what he believes to be truly valuable, the agent overturns the bigoted views he has been socialized into, and hence brings his motivations in line with what he takes to be authoritative.

2. Containment. The agent through thinking comes to realize that he does have desires of this sort and that they are ineradicable, and still thinks that the evaluative framework he has been socialized into is sound. He then proposes to treat himself as a recovering alcoholic might; trying to keep his recalcitrant desires under control through reason and making every attempt not to expose himself to temptation.

3. Hypocrisy. By an artful act of self-separation, the agent is able to have his cake and eat it. So long as the left hand does not know what the right hand is doing (or hides from itself the fact that it does), he can contrive a modus vivendi, though, of course he will be at a constant risk of the disparity between his two sides being brought to light.

4. Repression. The agent is aware in some sense that he is attracted to men. But he is unable to countenance he might be such an evil person. So he goes into overdrive, trying to prove to himself and to everyone else – he is terrified that they might guess his guilty secret – that he is definitely not gay.

Scenarios 1 and 2 are straightforwardly compatible with Taylor's understanding of the self. But 3 and 4 are more problematic.

The lesson to be learnt from scenario 3 (hypocrisy) is that the commitments which Taylor thinks define one's ipseity need not be operative in the context of action; they can be bracketed by the agent and ignored.

We could make things more difficult for Taylor's account of the self by proposing a more radical example: one in which a person, by living hypocritically and being recognized by others in a community for acting in this way, undermines the status of this way of living as hypocritical *for them* because they loosen the hold of the 'higher' set of values on them.

Consider a case where someone becomes a tyrant. Let us suppose that the initial act of violence by which they ascend the throne is done against their 'higher' self. Once they ascend the throne they continue to act against this higher self, and are buoyed up in the immorality of so doing by the flattering of their resident toadies. Before long, anyone who dares to speak out – to suggest that the actions coming from the tyrant's lower self are anything but fine – is swiftly despatched. It seems to me possible that such a process could continue until the higher self is completely destroyed (or, at the very least of no importance in defining where they stand) and we will be left only with the arbitrary exercise of power, validated as 'good' by a resident claque. It seems far from clear that such a person would necessarily suffer an identity crisis.

The lesson to be learnt from scenario 4 is that the self need not be held together positively, by a vision of a good, but can also be held together negatively – defining itself by what it is against, what it hates or resents. This target of resentment may even be another part of itself that it hides from itself. The verbalizations, the posturings through which people try to convey to the world and to themselves can just as well serve to mask desires and values that the person cannot admit to.

None of this speaks against Taylor model as an ideal: much of my criticism here turns on what one would consider to be included in the concept of 'undamaged human personhood'. Taylor could merely bite the bullet and exclude the tyrant or the extremely repressed from this; alternatively he could avail himself of a very general non-moral model of psychic health (as I have implicitly done above) in order to ground his ethics. In this case, the emphasis would not be on the *universality* of strong evaluation, but on articulacy about one's value as a condition of psychic health.

Smith reminds us, the transcendental argument (if successful)

> rules out contingency only at a very general level. It claims to show that there is a non-contingent relationship between self-interpretation and an orientation to the good, but it does not say anything about the content of the good. *That* a self-interpreting animal is non-contingently oriented against a background framework of strong evaluation does nothing to contradict the contingency of *how* a person is so oriented. (1997, p. 40)

5.5 *The argument for moral universalism*

I noted above that there are two apparent strategies for arguing from strong evaluation to moral universalism. Taylor does not explicitly commit himself directly to either at this point; rather he prefers instead to start from the pragmatics of ethical argumentation. Taylor here suggests that all ethical argumentation must be *ad hominem* – that is, addressed to people with particular conceptions of the good and work through conceptions of

157

the good to get to the correct answer. This follows quite directly from his position on the priority of the good to the right; if the right is always an attempt to articulate the good in rules, then clearly in ethical discussion we must start from what our opponent takes to be good.

More revealingly, Taylor suggests that we must take a certain frame of mind into practical argumentation:

> [P]ractical argument starts off on the basis that my opponent already shares at least some of the fundamental dispositions towards good and right which guide me. The error comes from confusion, unclarity, or an unwillingness to face some of what he cannot lucidly repudiate; and reasoning aims to show up this error. Changing someone's moral view by reasoning is at the same time increasing his self-clarity and self-understanding. (1993, p. 210)

What is the basis of this assumption? Is it just a regulative ideal; a certain attitude that makes argument possible in the first place? Or is it a substantive thesis, that, my opponent deep down really shares the same values as me? The first reading would turn on a transcendental argument about the conditions under which practical argument is possible; the second would turn on a substantive theory of human nature, namely that there is a core of good at the centre of each person that represents a bridgehead to be advanced from.

The following passage suggests that Taylor is committed to the more substantive thesis; he is prepared to take for granted a certain vision of human life as important:

> Do we really face people who quite lucidly reject the very principle of the inviolability of human life? In fact, this does not seem to be the case. Intellectual positions put forward to justify behaviour like that of the Nazis – to the extent that any of their ravings justify this appellation at all – never attack the ban on murder of conspecifics frontally. They are always full of special pleadings: for example that their targets are not really of the same species, or that they have committed truly terrible crimes which call for

retaliation ... The fact that these terrible negations of civilised morality depend so much on special pleading, and of a particularly mad and irrational sort, rather suggests that there are limits beyond which *rational* challenges to morality have a great trouble going. (1993, p.209)

Taylor concludes from this that 'The task [required] of reasoning, then, is not to disprove some radically opposed first premiss (e.g. killing people is no problem), but rather to show how the policy is unconscionable on premisses which both sides accept, and cannot but accept. In this case, its job is to show up the special pleas'. (1993, p. 209)

This is true as far as it goes, but when we recall Taylor's conception of normativity, we can quickly see that there is a severe problem here: Taylor argued i) that the right derives its normativity from the good, and ii) that good is inseparable from its ability to move us by intimations of what is higher or more worthy.

Taylor is here imagining that the dispute with the Nazi must take place here at the level of the *right*: the assumption is that the Nazi shares the same underlying sense of the good, but interposes some special pleadings, so that he can come out with his favoured account of the right. But this seems to discount the possibility of the difference occurring at the level of the *good*: that is, that a Nazi account of the right might be a perspicuous articulation of a totalitarian account of the good. Indeed, I think we can discern in the best fascist rhetoric a distinctive account of the good that could ground a Nazi account of the right in a way wholly consistent with Taylor's account of the relationship between the right and the good. Such a rhetoric, *pace* Taylor, does not attempt to justify rules or exclusions through special pleas, but to present us with a vision of life that inspires us and leaves us aching to realize it.

Ernst Jünger, in a book that proved to be one of the major spiritual sources of Nazism, tries to turn us on to the supreme worth of the Fatherland. The

159

following thoughts (he says) came to him as he was lying injured in hospital, wounded after his many heroics, as the war came to an end.

> And almost without any thought of mine, the idea of the Fatherland had been distilled from all these afflictions in a clearer and brighter essence. That was the final winnings in a game on which so often all had been staked: the nation was no longer for me an empty thought veiled in symbols; and how could it have been otherwise when I had seen so many die for its sake, and been schooled myself to stake my life for its credit every minute, day and night, without a thought? And so, strange as it may sound, I learned from this very four years' schooling in force and in all the extravagance of material warfare that life has no depth of meaning except when it is pledged for an ideal, and that there are ideals in comparison with which the life of an individual and even of a people have no weight...

> To-day we cannot understand the martyrs who threw themselves into the arena in a transport that lifted them even before their deaths beyond humanity, beyond every phase of pain and fear. Their faith no longer exercises a compelling force. When once it is no longer possible to understand how a man gives his life for his country – and the time will come – then all is over with that faith also, and the idea of the Fatherland is dead; and then, perhaps, we shall be envied, as we envy the saints their inward and irresistible strength. For all these great and solemn ideas bloom from a feeling that dwells in the blood and that cannot be forced. In the cold light of reason everything alike is a matter of expedience and sinks to the paltry and mean. It was our luck to live in the invisible rays of a feeling that filled the heart, and of this inestimable treasure we can never be deprived. (Jünger 1929, pp. 316-7)

The problem for Taylor is this: Jünger will agree on the importance of all human life[119] but will argue, *on the level of the good,* that the value represented by the Fatherland is immeasurably more compelling. While others might be able to try to argue against such a position by putting forward an account of the right against which to judge such an account of

---

[119] For example, his preface to the English edition expresses 'sincere admiration' (p.xiii) of the British forces.

the good as wrong, deluded and dangerous, Taylor can do nothing of this sort.

Instead he must hope that it can be shown that there exists a 'way of experience' through which someone like Jünger could be brought from his position to ours. I frankly doubt that this could be done. What if this cannot be done? If there is no such way of experience for each and every person, then it seems that we cannot establish morality on the basis of the subjectivist account of strong evaluation.

This would mean that Taylor would have to adopt the second strategy; that of arguing for an objectivist ontology for strong evaluation. The problem is, I do not see how this strategy could work: Taylor ties the good to motivation constitutively, so it is hard to see (given his assumptions about normativity) how an objective good could be normative for me if I proved motivationally immune to it. Hence the objectivist account must collapse into the subjectivist one.

Taylor is aware of this problem, but believes that its import runs in a different direction – not to the inadequacy of his account of ethical thinking, but to a deep problem with our whole conception of morality. In *Sources of the self,* he suggests that the crisis for moral universalism lies not in the fact that we find ideas such as universal human rights incorrect, but that we have lost touch with the motivational sources that originally powered the ideals of universal morality. The problem is not so much in our inability to convince individual people of the merits of morality, but that we ourselves have lost touch with the goods that underlie our commitments to morality, so that our account of the right is no longer anchored in the good:

> High standards need strong sources. This is because there is something morally corrupting, even dangerous, in sustaining the demand simply on the feeling of undischarged obligation, on guilt, or its obverse, self-

satisfaction. Hypocrisy is not the only negative consequence. Morality as benevolence on demand breeds self-condemnation for those who fall short and a depreciation of the impulses to self-fulfilment, seen as so many obstacles raised by egoism to meeting our standard...

The question which arises from all this is whether we are not living beyond our moral means in continuing allegiance to our standards of justice and benevolence. Do we have ways of seeing-good which are still credible to us, which are powerful enough to sustain these standards? If not, it would be both more honest and more prudent to moderate them. (1989, pp. 516-7)

In counterpoint to this sharply stated choice, Taylor suggests that it may only be by a turn to theism that we can find a moral source sufficient to power the ideas of moral universalism. I must at this point be honest: I cannot follow Taylor in this proposed return to theism. If morality can be shown to require God, then (from my perspective) the most honest thing would be to give up on universalistic morality.

5.6 Conclusion

We spent the first half of this chapter demonstrating that the considerations that Taylor invokes (his critiques of the representationalist conception of knowledge, the idea of a purely procedural ethics, his descriptive phenomenology, and his principle of the best account) do not necessitate his conclusion that ethical thinking must consist in articulating the substantive horizon of strong evaluations that we already committed to.

The second half worked within Taylor's conception of ethical thinking to demonstrate the difficulties that would be involved in defending morality on this basis. I raised some difficulties with the transcendental argument he gives for taking strong evaluation to be a necessary condition for human agency, which amounted to pointing out that psychological economies can be much more complex than is implied by Taylor's model; that a person's self need not be solely held together by strong evaluations.

But there are deeper, structural reasons for finding Taylor's account inadequate: he elucidates the good *in terms of* a certain sort of motivation, but then must be committed to the existence of a 'way of experience' from any set of strong evaluations to those involved in moral virtue. It seems unlikely that we will be able to do this. What is more Taylor's account of moral normativity entails that we must reinflate our account of moral right with the breath of the good; it seems highly unlikely that we will be able to do this without God.

The account of morality I shall give in part three is strongly influenced by what we have learnt from this study of Taylor's moral philosophy, but it takes a path that Taylor himself does not take. Like Taylor's, it is an ethics which takes seriously the relationship between selfhood, moral identity and narrative. It differs fundamentally, however, in that where Taylor's account takes the *vertical* relationship between the self and something that is experienced as *incomparably higher* as foundational to his ethical thinking, I take the *horizontal* relationship of reciprocal recognition to be foundational to mine.

In addition, we make use of the fact that the normativity of morality need not be reducible to that of the good: in this case, morality can unproblematically go beyond our ability to be deeply moved by it without thereby undermining its own normativity. In the final chapter I try to set out an account of moral normativity, that is grounded (as Taylor's arguments require) in the human good, but which is not directly answerable to this human good. This way we will be able to defend moral normativity within a secular framework.

In so doing, I hope that we will be able to construct an account of morality that is both satisfying to ourselves and resistant to the sorts of criticism that Nietzsche would want to make of it.

PART THREE:

A PRAGMATIST

REINTERPRETATION OF

MORALITY

MAP OF PART THREE

In part three, I fashion a pragmatist reply to Nietzsche's claims. The account of morality is split over three chapters.

Chapter six, *Recognition and Human Life,* is meant as a riposte to Nietzsche's underlying theory of human nature with its glorification of the will to dominate. It argues that *mutual recognition* has a central place in human life, and that the will to dominate only arises out of a failure in the process of recognition. It is quite unnecessary for human relationships to resolve themselves into the pattern of dominator and submissive: there is another alternative, mutual recognition, and it is only this alternative that can solve the problem that domination attempts to solve.

Chapter seven, *Dignity and Vulnerability,* is concerned with the dignity of human beings. Here I argue that Nietzsche in fact has much in common with the tradition of morality: both laud hardness and self-sufficiency as the appropriate response to life, and equate the dignity of human beings with this hardness. For both, vulnerability is a form of weakness that should be extirpated. Against this, I return to an alternative tradition of thinking that dovetails well with the account of recognition, which places a positive value on human vulnerability and understands the dignity of human beings in terms of their courage *in the face of* the riskiness and arbitrariness of life.

Chapter eight, *Reconstructing Universality,* completes the reconstruction of morality by giving an account of moral rules, and by extending and clarifying what I have so far said about dignity.

The account of moral rules does three things. First, it explains what the regulation of life through moral rules is supposed to accomplish, namely

the continuing viability of morality as a project, and the equal respect of each person under rules that all can agree to. Second, it articulates why one might find this goal worth striving towards. Third, it gives a theory of argumentation which we can use to determine which norms to take to be morally valid.

The account of dignity makes a distinction between a bloodless and a full-blooded conception of the dignity of human beings. The bloodless version is all that is presupposed by the project of constructing moral rules, and amounts to no more than considering a person to be a part of the moral community. The full-blooded version takes its cue from the conception of the hero's dignity, but transforms this so that we can universalise it. It requires far more of us, but it promises a more effective affirmation of human life.

Last, I give a vision of moral life, in which moral rules and this full-blooded conception of human beings are combined.

Chapter Six

Recognition and Human Life

6.1 *Introduction*

This chapter argues that Nietzsche is mistaken in his basic account of human nature in the place he gives to the will to dominate, because what he isolates and lauds is in fact just a failure in the search for recognition. It is quite unnecessary for human relationships to resolve themselves into the pattern of dominator and submissive: there is another alternative, mutual recognition, and it is only this alternative that can solve the problem that both domination and submission attempt to solve.

For Nietzsche, life is essentially will to power, that is the will to dominate, to be master:

> [L]ife is *essentially* appropriation, injury, overpowering of the strange and weaker, suppression, severity, imposition of one's own forms, incorporation and, at the least and mildest, exploitation – but why should one always have to employ precisely those words which have from of old been stamped with a slanderous intention? … 'Exploitation' does not pertain to a corrupt or imperfect or primitive society: it pertains to the *essence* of the living thing as a fundamental organic function, it is a consequence of the intrinsic will to power which is precisely the will of life. (1886, §259)

The will to power can be turned inward as well as outward (or it can be turned in both ways simultaneously): when turned inward, the will to mastery is applied to the self, and we end up with a form of asceticism. When turned outwards we have the will to dominate others. We shall not be concerned with the will to power as turned inwards in this chapter: insofar as the will to power is focused inwards as asceticism a good case can be made for its usefulness.[120]

---

[120] See for example, Nehamas (1985). The account of how it is possible to believe in the dignity of all human beings that I shall give in chapter eight makes use of some ascetic practices of self-

In line with Nietzsche's technique, I shall not attempt to argue against this interpretation so much as to offer another interpretation which subsumes it, and which makes Nietzsche's interpretation look biased and partial by comparison.

Compared to Nietzsche's account this interpretation has the following advantages: i) it incorporates a lot more empirical observation of the behaviour of human beings and integrates the insights of many more perspectives: philosophy, child psychology and psychoanalysis. ii) It tells a developmental story of how human beings come to be what they are; iii) it casts light on the instances of exploitation and the will to dominate that Nietzsche wishes to illuminate but also sheds light on other things that Nietzsche's theory fails to explain.

This theory places not the will to dominate, but the *struggle for recognition* at the centre of human life, and reinterprets the will to dominate as a result of the failure of the process of recognition. Recognition is helpfully defined by Jessica Benjamin as 'that response from the other which makes meaningful the feelings, intentions and actions of the self. It allows the self to realize its agency and authorship in a tangible way… Recognition is so central to human existence as to often escape notice; or rather, it appears to us in so many guises that it is seldom grasped as one overarching concept. There are any number of near-synonyms for it: to recognize is to affirm, validate, acknowledge, know, accept, understand, empathize, take in, tolerate, appreciate, see, identify with, find familiar, … love' (1988, pp. 12, 15-16)

I shall argue that recognition is a vital human need because it is *only* through recognition that an individual can come to feel real to herself – that is to feel validated as a person and to find her own actions and life meaningful.

---

modification, though the asceticism is unNietzschean in tone, aiming as it does at the expansion of curiosity and love for others, rather than a reshaping of the self in the direction of hardness through a self-directed cruelty.

The theory that I shall present takes its cue from Hegel's early writing on recognition, up to and including the *Phenomenology of Spirit.* Hegel's thought is at this point working in two directions at once.

First, he argues that an analysis of the romantic love relationship gives us an insight into the basic core of ethical life. Each person can only fully come to be themselves *indirectly* through the other's validation of them:

> Each one is identical to the other precisely owing to that in virtue of which each is opposed to the other; the other, that through which the other exists for one, is oneself.[121]

Habermas helpfully glosses this as follows:

> In a symmetrical relation the point of mutual recognition is that the two persons involved seem to sacrifice their independence; but in fact each gains a new kind of independence by coming to recognise, in the mirror of the eyes of the other person, who he or she is. Both become for themselves the kind of characters they mutually attribute to each other. Both gain awareness of their individuality by seeing their own images reflected in the dense and deep exchange of an interpersonal relation. (1999, p.140)

The other aspect, which we see in the *Phenomenology of Spirit* §178-196 is the argument that there is something inherently unstable in the attempt to get this recognition through *force,* because we can only be recognised by an other whom we recognise as a person in his or her own right. One comes to find one's life unreal and hollowed out to the extent that is unable to recognise others as persons in their own right.

Hegel dramatizes this through the famous 'master-slave' dialectic. Hegel begins by supposing two self-consciousnesses, both of whom want recognition from the other. If they fight and one is killed, then even the victor is no nearer to receiving the recognition he craves: for you cannot receive recognition from a dead man. Suppose then instead the conflict

---

[121] Hegel (1983, p.107). See also *Phenomenology of Spirit* §184: "Each is for the other the middle term, through which each mediates itself with itself and unites with itself; and each is for itself, and for the other, and immediate being on its own account, which at the same time is such only through this mediation. The *recognize* themselves as *mutually recognizing* one another."

falls short of death, and we are left with a situation in which the stronger becomes the master and the weaker his slave. But this situation is little better, for by enslaving the other, and treating him as a thing, he fails to confirm his reality in the way he wanted to.[122] Interestingly, the slave, meanwhile is better off at least to the extent that the master for him is *real* and is able to form expressive relations with the things he makes through his work.

The overall thesis is ably summarised by Honneth:

> [A]n individual that does not recognize its partner to interaction to be a certain type of person is also unable to experience itself completely or without restriction as that type of person. The implication of this for the relationship of recognition can only be that an obligation to reciprocity is, to a certain extent, built into such relations, an obligation that requires but does not force subjects to recognize one another in a certain way: if I do not recognize my partner to interaction as a certain type of person, his reactions cannot give me the sense that I am recognized as the same type of person, since I thereby deny him precisely the characteristics and capacities with regard to which I want to feel myself affirmed by him. (Honneth, 1995, pp. 37-8)

6.2 *Winnicott's theory of child development*

But considered in themselves, Hegel's arguments seem just as unsupported as Nietzsche's arguments for the will to power: we need an empirically grounded theory that will buttress these claims.

Donald Winnicott's work provides the basis of such a theory. However, Winnicott's theory comes out of a very different framework from the one in which we have hitherto been working: Winnicott was a paediatrician and

---

[122] *Phenomenology of Spirit* §192. "In this recognition the unessential consciousness is for the lord the object, which constitutes the *truth* of his certainty of himself. But it is clear that this object does not correspond to its Notion, but rather that the object in which the lord has achieved his lordship has in reality turned out to be something quite different from an independent consciousness. What now really confronts him is not an independent consciousness, but a dependent one. He is, therefore, not certain of *being-for-self* as the truth of himself. On the contrary, his truth is in reality the unessential consciousness of its unessential action."

child psychotherapist, who positioned his work relative to the body of work in psychoanalysis, and especially the work of Melanie Klein.

What is required therefore is a translation of the core of Winnicott's theory into a medium that is usable by us. Luckily, much of the hard work has already been done by Jessica Benjamin and Axel Honneth, and I shall use their analyses where appropriate.[123]

Winnicott's theory aims to explain how a new born infant, unable as yet to distinguish between inner and outer, what belongs to self and what to the world, becomes a *subject* capable living in an objective world and relating to other subjects as subjects, and the reasons why this process sometimes fails or gets stalled. His answer, is that what makes the difference is the quality of the relationship between the infant and the primary care giver[124] and the progress in styles of recognition.

Where this care fails (in Winnicott's terminology, when the mother is not *good enough*) various forms of psychopathology predictably result. The nature of these psychopathologies will depend both on the stage in development at which the mother fails to be good enough and the way in which this inadequacy is manifested.[125]

Winnicott's theory envisages three basic stages: first *absolute dependency*, next *relating to subjective objects* and finally *use of objects*.[126]

---

[123] I have also found Hundert (1989) helpful.

[124] Or, in Winnicott's now slightly dated and sexist sounding terminology, the 'mother'. 'Primary care giver' is however a cumbersome phrase, and I shall continue to use the word mother, with the proviso that we should be clear that it is the role of the primary care giver in infancy that Winnicott is theorizing, and, of course this role need not in fact be played by the baby's actual mother.

[125] Note, however that the 'good enough mother' must be the norm and the not good enough mother less frequent: it is difficult to see that the human race could have progressed this far had good enough mothering been rare. As Winnicott puts it, "it is important that the function of the mother should be understood. This function is by no means a recent development, belonging to civilisation or to sophistication or to intellectual understanding. No theory is acceptable that does not allow for the fact that mothers have always performed this function well enough." (1960a, pp.147-8)

[126] A warning: as will become apparent Winnicott's terminology is quite counter-intuitive. As Benjamin explains: "'using' here does not mean instrumentalizing or demeaning, but being able to creatively benefit from another person; it refers to the experience of "shared reality" in which "the object's independent existence" is vital. "Relating" refers to the experience of "the subject as isolate," in which the object is merely a "phenomenon of the subject."' Benjamin (1988, p. 37)

Corresponding to these three stages is a gradually increased capacity of the infant to be independent and a decreased fragility in the infant's self. At first dependence is total; the self is easily ruptured and so the mother needs to be totally attuned to the infant's needs if development is to occur: the infant needs an environment that allows it to just be, without having to *react* to the outside world.[127]

Luckily, correlative to this basic need of the baby is the good enough mother's Primary Maternal Preoccupation: that is, a high degree of identification with the infant that develops through the pregnancy, which enables her to 'know what the infant feels like, and so … provide almost exactly what the infant needs at the beginning, which is a *live adaptation to the infant's needs.*' (1960b, p.54) It also makes the mother vulnerable and dependent on the infant, because she experiences its cries as a failure, as a lack of sensitivity on her part.[128]

What happens next depends on whether the mother is good enough or not.[129] The good-enough mother will first help the infant's self to grow by implementing his desires, and do this so well that the infant will have a brief experience of omnipotence: that is, the infant will experience its desires in such a way that it seems they automatically bring about their

---

[127] As Winnicott puts it in an important early paper, 'health in the early development of the individual entails continuity of being. The early psyche-soma proceeds along a certain line of development provided its continuity of being is not disturbed; in other words for the healthy development of the early psyche-soma there is a need for a perfect environment. At first the need is absolute. The perfect environment is one which actively adapts to the need of the newly formed psyche-soma, that which we as observers know to be the infant from the start. An environment is bad because by failure to adapt it becomes an impingement to which the psyche-soma (i.e. the infant) must react. This reacting disturbs the continuity of the going on being of the infant.' (1949, p.245)

[128] Benjamin 1988, p.28: "Just as the baby's positive response can make the mother feel affirmed in her being, the baby's unresponsiveness can amount to a terrible destruction of her self-confidence as a mother. The mother who jiggles, pokes, looms, and shouts 'look at me' to her unresponsive baby creates a negative cycle of recognition out of her own despair at not being recognized. Here in the earliest social interaction we see how the search for recognition can become a power struggle: how assertion becomes aggression."

[129] "So much difference exists between the beginning of a baby whose mother can perform this function well enough and that of a baby whose mother cannot do this well enough that there is no value whatever in describing babies in the earliest stages except in relation to the mother's functioning." (1962, p.57)

fulfilment.[130] She then gradually de-adapts to the child (with the fading of Primary Maternal Preoccupation) in such a way that he or she gradually comes face to face with elements of reality.

The mother who is not good enough, however, is 'not able to implement the infant's omnipotence, and so she repeatedly fails to meet the infant gesture; instead she substitutes her own gesture which is to be given sense by the compliance of the infant. This compliance on the part of the infant is the earliest stage of the False Self, and belongs to the mother's inability to sense her infant's needs.' (1960a, p. 145)[131]

As the good-enough mother gradually de-adapts to the infant, the infant comes face to face for the first time with the fact that the mother is not just a subjective object, under the control of his fantasying (this is the illusion given by the good-enough mother's responsiveness to the infant's needs).

This situation creates a challenge: in Winnicott's terms, the infant must go from *object relating* to the *use of an object.* Winnicott argues that what must occur is that the subject places the object outside the area of his omnipotent control, that is sees the object as independent from him.

On Winnicott's account, the infant's destructiveness is of prime importance here: it is only by the infant repeatedly attempting to destroy the object, and the object surviving this attack that the quality of externality, of the reality of the external world is created.[132]

This is Winnicott's theoretical innovation, and it is important to stress it. Whereas in Freudian theory, (and in any theory such as Nietzsche's which

---

[130] "The good-enough mother meets the omnipotence of the infant and to some extent makes sense of it. She does this repeatedly. A True Self begins to have life, through the strength given to the infant's weak ego by the mother's implementation of the infant's omnipotent expressions." (1960a, p.145); see also (1962, p.57).

[131] The development of an extensive false self system of this sort is an important factor predisposing the infant to schizophrenia in later life. See R D Laing (1960, chapter six) on this.

[132] As Winnicott puts it: "This thing that there is in between relating and use is the subject's placing of the object outside the area of the subject's omnipotent control; that is, the subject's perception of the object as an external phenomenon, not as a projective entity, in fact recognition of it as an entity in its own right." (1969, p.89)

gives pride of place to the will to dominate) the incursion of the hardness of reality is seen as an incursion and a source frustration, on Winnicott's account, the survival of destructiveness is a source of joy and reality for the individual.[133] As Benjamin puts it, 'Beyond the sensible ego's bowing to reality is the joy in the other's survival and the recognition of shared reality.' (1988, pp. 40-1)

Destruction can fail in one of two ways: either the object does not survive or the object retaliates. Where the object does not survive, we have the proto-type of what will reappear in the adult as sadism, or the will to dominate: where the object retaliates, we have the proto-type of what will reappear in the adult as masochism, or the will to be dominated.

The mother must pursue a delicate balancing act: setting limits for the child allowing the space for his true self to develop:

> So if the mother sets no limits for the child, if she obliterates herself and her own interests and allows herself to be wholly controlled, then she ceases to be a viable other for him. She is destroyed and not just in fantasy. If she retaliates, attempting to break his will, believing that any compromise will "spoil" him, she will also inculcate the idea that there is room for only one ego in any relationship – he must obliterate his for now, and hope to get it back, with a vengeance later. Only through the other's survival can the subject move beyond the realm of submission and retaliation to a realm of mutual respect. (Benjamin, 1988, p.39)

When the object does not survive, the child continues to attack, seeking a boundary for his anger and destructiveness. But for him, unfortunately 'the real object, the one who cannot be destroyed, never comes into view. For him, assertion and agency are not integrated in the context of mutuality and respect for the other but in the context of control and retaliation. The sadist-child is *cognitively* aware of the difference between self and other,

---

[133] "The assumption is always there, in orthodox theory, that aggression is reactive to the encounter with the reality principle, whereas here it is the destructive drive that creates the quality of externality. This is central in the structure of my argument." (1969, p.93)

but *emotionally* this awareness is hollow and does not counteract the desire to control the other.' (Benjamin 1988, p.20)

This success of domination has results far from desirable, even for the dominator: for the child experiences the parent's caving in as 'flying off into space – he finds no limits, no otherness… When the other crumbles under my act, then my act seems to drop off the edge of the world into emptiness, and I feel that I will soon follow. In this void begins the loss of tension or boundaries, a by-product of losing the other.' (1988, pp. 70-1)

When the mother retaliates, the situation is scarcely more enviable: the child does not have the chance to discover his own sense of agency and his own true self; 'he has not experienced his own impulses and acts as his own, arising without direction from outside.' (Benjamin 1988, p.72) He 'despairs of ever holding the attention or winning the recognition of the other, of being securely held in the other's mind.' (Benjamin 1988, p.72)

6.3 *Recognition and domination in adult relationships*

It is time to step away from Winnicott's theory as such and to apply what we have learnt to the understanding of adult relationships.

The theory we are pushing takes it that the most valuable forms of intersubjectivity such as true friendship and erotic love result from an experience of temporary merging between those who have developed what Winnicott calls the *capacity to be alone*.[134] The capacity to be alone is a feeling of security and confidence that allows a person to *relax* when in their own company. In such relationships, each *recognises* the other as a separate capable subject, and this recognition opens up a communal space in which both experience *being with* the other. As Benjamin puts it,

---

[134] Winnicott 1958b.

One of the most important insights of intersubjective theory is that sameness and difference exist simultaneously in mutual recognition. … Experiences of 'being with' are predicated on a continually evolving awareness, on a sense of intimacy felt as occurring between 'the *two* of us.' The fact that self and other are not merged is precisely what makes experiences of merging have such high emotional impact. The externality of the other makes one feel one is truly being 'fed,' getting nourishment from the outside, rather than supplying everything for oneself. (1988, p.47)[135]

Friendship and love are thus shown to be quite fragile affairs requiring a continual play between the participants, *both of whom must* be comfortable both to be alone and to be merged with the other. When these exacting conditions are not met we have a relationship of domination and submission rather than a 'communicative arc suspended between the experience of being able to be alone and the experience of being merged' (Honneth, 1995, p.105)

It is easy to see that such experiences open up each individual to the possibility of harm: they are entrusting what is most precious to them to a common space created with another: such a move is risky. It runs the risks both of bad luck – for example of the other dying or becoming incapacitated: or worse still of betrayal. But there is no way of evading these risks without sacrificing the experiences that recognition opens up.

Jessica Benjamin argues that the sado-masochistic relationship is key to understanding how the failure of destruction manifests itself in later life. Erotic domination attempts to compensate for the *failure* of the process of destruction. But it does so in an ultimately futile manner:

---

[135] In Honneth's more jargon ridden formulation, "We can then proceed from the hypothesis that all love relationships are driven by the unconscious recollection of the original experience of merging that characterised the first months of life for 'mother' and child. … Of course, this desire for merging can only become a feeling of love once, in the unavoidable experience of separation, it has been disappointed in such a way that it henceforth includes the recognition of the other as an independent person. Only a refracted symbiosis enables the emergence of a productive interpersonal balance between the boundary-establishment and boundary-dissolution that, for Winnicott, belongs to the structure of a relationship that has matured through mutual disillusionment." (1995, p.105)

176

> Domination presumes a subject already caught in omnipotence, unable to make 'live' contact with outside reality, to experience the other person's subjectivity. But this apparent first cause is itself the result of an earlier breakdown between self and other – which though pervasive, is not inevitable. Insofar as domination is an alienated form of differentiation, an effort to recreate tension through distance, idealization, and objectification, it is destined to repeat the original breakdown unless and until the other makes a difference. (1988, p.68)

The problem is exactly the one that we saw with the master-slave relationship in Hegel.[136] Both sadism and masochism have to keep on raising the stakes in an attempt to find the elusive recognition that they have unfortunately ruled themselves out from at the beginning. Such a spiral can only end in disaster, if neither finds a way out:

> Eventually the other's unreality becomes too powerful; the sadist is in danger of becoming the will-less thing he consumes unless he separates himself completely. And the masochist increasingly feels that she does not exist, that she is without will or desire, that she has no life apart from the other. (1988, p.65)

What I hope to have established in this account is that the universal dynamics of human life, far from favouring the will to dominate, in fact ensure that the will to dominate will regularly and reliably fail to provide the sort of satisfaction that the dominator is looking for. The dominator seeks a sense of reality, of someone else there who can *make a difference,* but his attempt to dominate in fact undermines the possibility of this occurring. Furthermore, this same account also suggests that the most fulfilling human experiences will be those of reciprocal exposedness, where one person feels he is getting fed from the outside: we experience

---

[136] "We might call this the dialectic of control: If I completely control the other, then the other ceases to exist, and if the other completely controls me, then I cease to exist. A condition of our own independent existence is recognizing the other. True independence means sustaining the essential tension of these contradictory impulses; that is, both asserting the self and recognizing the other. Domination is the consequence of refusing this condition." (1988, p.53)

our lives as valuable to the extent that we can open up and enjoy mutual experiences with others.

The next chapter pushes these insights further, and rethinks the dignity of human beings.

Chapter Seven
Dignity and Vulnerability

7.1 *Introduction*

This chapter argues against *both* Nietzsche *and* the tradition of morality: both laud *hardness and self-sufficiency* as the appropriate response to life, and equate the dignity of human beings with this hardness. For both, vulnerability is a form of weakness that should be extirpated. Against this, I return to an alternative tradition of thinking that dovetails well with the account of recognition we put forward in the previous chapter, which places a positive value on human vulnerability and understands the dignity of human beings in terms of their courage *in the face of* the riskiness and arbitrariness of life. On this account, the hard, self-sufficient person has cut themselves off from some of the deepest sources of value: their self-sufficiency comes to seem less heroic and more like a strange combination of wishful thinking and self protection.

The mainstream of the tradition of morality concentrates on the *agent* to the exclusion of the *patient* of action: it is far less interested in the *effects* of various forms of immorality (violence or betrayal for instance) on the unfortunate recipient of these violations than it is in the promulgation of universal rules and the accompanying condemnation of the agent who transgresses them.[137]

This lack of balance ought to be a little surprising, given the centrality of the concept of reciprocity to morality: in chapter one we saw that the golden rule by its very nature focuses attention on the passive recipient of the action and hence, one would assume, should make the agent aware of

---

[137] The consequentialist tradition of ethical thinking has of course been more sensitive to the suffering of the vulnerable.

the possibility of the misuse of his power and its potentially devastating effects on the other.[138]

I believe that this state of affairs is best explained by the fact that there is another tradition of ethical thinking that has become entwined into morality, and has done so through morality's understanding of the dignity of human beings.

This other tradition of ethical thinking (which, following Martha Nussbaum, I shall call the anti-pity tradition) is concerned to claim that human goodness is self-sufficient and hence to claim that each person always has within his[139] power sufficient resources to lead a flourishing life. According to this tradition, human beings are *not* basically needy and vulnerable creatures: such weaknesses are a result of a lack of self-mastery which is ultimately to be traced back to a lack of virtue.

This anti-pity tradition is relatively uninterested in the effects of the violation of a human subject because it sees such violations as *as such* unable to break up the goodness (flourishing) of the virtuous person's life. Insofar as the goodness (flourishing) of a person's life is capable of being broken up from the outside by violation that person lacks a necessary part of virtue. Violation, and in general serious suffering is thus, according to this tradition, uninteresting for ethical theory because it only really affects people who lack virtue: that is it only affects people who we already know should be other than they currently are.

When this tradition is entwined into morality, this tendency to play down or efface the effects of violation is, unfortunately, exacerbated. This is perhaps a little surprising: one might imagine that morality's insistence on

---

[138] Especially in Rabbi Hillel's formulation: "What is hateful to you, do not do to your neighbour." Ricoeur (1992, p.219) underscores this point 'The most remarkable thing, however, in the formulation of this rule is that the reciprocity demanded stands out against the background of the presupposition of an initial dissymmetry between the protagonists of the action – a dissymmetry that places one in the position of agent and the other in that of patient. ... it is upon this dissymmetry that all the maleficent offshoots of interaction, beginning with influence and culminating in murder, will be grafted.'.

[139] 'His' is used deliberately here to hint at the latent sexism of such theories.

the dignity of human beings would lead to an *increased* awareness of the results of violating, insulting or otherwise failing to respect that dignity.

But this is not so. The mainstream of morality (following the Stoics, and especially Kant) has understood the dignity of a human being to *be* that person's capacity for virtue.[140] This entails that should some violation break up a person's capacity for virtue, it would *ipso facto* break up their dignity and thus their claim to respect from their fellow human beings.

This possibility is deeply problematic from the perspective of morality: for it would undermine the claim that *all human beings have a dignity that renders them worthy of respect,* and turn dignity into a *contingent* quality of a human being. What is more, it would imply that a person's claim to respect might be destroyed by another's unjust violation, whilst the one who had done the wrong would *retain* his claims to the respect of his fellow men.

The problem in short is this: the way the anti-pity tradition is entwined with morality, it becomes impossible to admit that someone could be broken by circumstances – say, by the murder of their child or the suicide of a lover – because their claim to respect (which morality knows they cannot lose) *requires* them to remain in control.

The tradition of morality attempts to avoid this unpalatable possibility by stipulating that no violation *could* break up a person's capacity for virtue. But this has the effect of skewing moral thinking: because no one can be rendered incapable by violation, thinking proceeds from the perspective of the capable agent and his duties.

---

[140] This capacity for virtue is understood in terms of *practical reason,* where practical reason is taken to be the ability to order one's life in accordance with the dictates of morality.

7.2 *The Pro-Pity tradition*

Just as we traced the tradition of morality back to its beginnings in Stoicism in chapter one, so we must trace the pro and anti pity traditions back to their beginnings. These too are in the thought of the Greeks.

Greek ethical thought before Socrates takes it for granted that human life is a chancy affair, and that if one is to live a flourishing life one needs *luck* as well as virtue. Happy lives can be torn apart by fate or by the malice or incompetence of others. This possibility is not something that can be avoided or warded off; rather it calls for humility, a creation, if you will of a community of sufferers, ready to respond with pity to others in respect of those things they fear for themselves.

Book 24 of the *Iliad* provides the best example of the heights to which this approach to ethical thinking may rise. Frail Priam, the King of Troy, aided by the god Hermes, arrives at the tent of Achilles. He has come to ransom the body of his favourite son Hector, whom Achilles has slain. He has driven through the night at great danger to himself. He enters unseen, takes Achilles in his arms and kisses his hands, 'those terrible, murderous hands, which had killed many of his sons,' and says

> "Think of your father, godlike Achilles, an old man like I am, at the cruel edge of old age. And it may be that he too is pressed by those who live around his home, and there is no-one to protect him from harm and destruction. But he can at least hear that you are alive, and feel joy in his heart, and look forward every day to seeing his dear son return from Troy. But my fate is utter misery – I fathered sons who were heroes in the broad land of Troy, and I tell you not one of them is left. ... Respect the gods, then Achilles, and have pity on me, remembering your own father. But I am yet more pitiable than he. I have endured to do what no other mortal man on earth has done – I have brought to my lips the hands of the man who killed my child."

> So he spoke, and he roused in Achilles the desire to weep for his father. He took the old man by the hand and gently

pushed him away. And the two of them began to weep in remembrance. Priam cried loud for murderous Hector, huddled at the feet of Achilles, and Achilles cried for his own father, and then again for Patroclus:[141] and the house was filled with the sound of their weeping. (Book 24, 482ff.)

It is important to note that it is through the emotion of pity that the mutual recognition of Priam and Achilles takes place. Through pity they come to see each other as equally human, both with symmetrical fears: Priam fearing for his sons and the destruction of his line; Achilles fearing for his father, growing old without a son to take care of him.

Aristotle's discussion of pity in the *Rhetoric* is the most helpful place to go if we wish to get a philosophical handle on this tradition of ethical thinking: he defines pity as a 'feeling of pain at an apparent evil, destructive or painful, which befalls one who does not deserve it, and which we might expect to befall ourselves or some friend of ours, and moreover to befall us soon.' (1385b12-14)

There are three elements to this definition of pity: first, the evil has to be serious – ('we do not go around pitying someone who has lost a toothbrush or a paperclip' (Nussbaum 1994, p.141); second, the evil cannot be the person's fault, or if it is their own fault, the evil must be far greater than would be appropriate given the fault;[142] and third that the pain of pity requires one to see befalling a similar evil as a real possibility for oneself.[143]

Martha Nussbaum argues that were this tradition to rise to the level of philosophical theorizing, it would posit three distinct advantages for its pity-oriented outlook: first, *pity expresses the truth about the human*

---

[141] Achilles' best friend, whom Hector had slain.

[142] As a side issue, Aristotle's theory of tragedy demands that the protagonist does not suffer for no reason at all, but rather that the suffering must be caused by a mistake (*hamartia*) though the punishment must be so extreme as to excite pity.

[143] 'What we fear for ourselves excites our pity when it happens to others.' (1386a27-8)

situation: we are needy incomplete creatures who are reliant on external circumstances for our flourishing.

Second, this truth holds regardless of how rich or powerful a person may be: everyone should bear this in mind, because 'one will act more appropriately if one remains aware at all times of the sort of being one is.' (1994, p.143) Third, because pity acknowledges the importance of external goods, and because pity requires one to admit that one may one day be in need oneself, pity 'works to ensure a more equitable distribution of goods.'(1994, p.143)

7.3 *The Anti-Pity tradition*

But Socrates, as we see him in the *Apology,* the *Crito* and *Phaedo* represents a very different model of life from Priam and Achilles. Faced with the prospect of being put to death unjustly, Socrates does not cry, get angry or violent: rather he remains preternaturally calm, insisting that he has not really been harmed: for to be really harmed he would have to have become unjust. For according to Socrates, the most flourishing life for a human being is one of virtue: everything else is of minor importance.

From Socrates' perspective, the whole pro-pity tradition is deeply flawed: its most basic error is that it assumes that externals such as money, power friends and children are important, when they are not. Pity is insulting to the pitied because it implies that the pitied person really needs these externals, when any virtuous person does not. Moreover it is demeaning to the one who pities, because it implies that she too needs externals, when it is clear that all that is required is virtue. Rather, no serious harm can befall the just person qua just person, for the only injury that would count as serious would be one that interfered with their virtue.

This tradition of anti-pity thinking that starts with Socrates is entwined with morality right from the beginning through the conception of the

dignity of human being that is present in Stoicism. Human dignity is thought to consist in reason, and reason to be the identical to the capacity for virtue.

If you understand human dignity as being grounded in *reason,* and reason to involve *mastery* over the emotions, then it is easy to get to the thought that responding appropriately to one's dignity as a human being is all about mastering the emotions, and hence that someone who is overcome by emotion (say at the ending of a love affair) thereby either diminishes their dignity or (if that is not possible) fails to act in a way that is respectful of that dignity. George Harris summarizes the essence of this tradition well:

> According to this tradition, when character either weakens or succumbs to life's trouble, it fails because it lacks the kind of strength ideal characters should have. To be sure, good people are almost always less than ideal. For this reason we understand that certain failures of strength are compatible with being a good person. Still we would be better were we to realize fully that which gives us our dignity, and if we did fully recognize our dignity we would have unlimited strength to cope with life's troubles whatever they might be. Were we good Kantians and more rational, were we good Christians and more faithful to God, and were we good Stoics and less attached to things external to our character, we would not be vulnerable to failures of strength. Being more dignified, we would be stronger. (Harris 1997, pp. 1-2)

Set against the background of this personality ideal, the travails and possible breakdown of the person who suffers the effects of immorality look like they occur because of a failure of those very qualities – hardness towards oneself and strength of character – that make the good person good. The unavoidable conclusion is that it is due to a failure on the part of the person who suffers that they are overcome by suffering. The natural conclusion to draw is that drawing attention to such suffering is much less important than telling people how to overcome it through their own

hardness. People are only incapacitated to the extent that they allow themselves to be: what is morally important is to focus on capability.

The combined effect of these decisions is that the *vulnerability* of human beings has tended to become hidden by a zealous affirmation of the inalienable capability of human beings to meet their moral demands. The person who thinks of him or herself as having been broken on the wheel of life is just being weak: it is in their power to stop being weak in this sense, and 'pull themselves together', and so their weakness is, ultimately, *morally culpable.*

I think that this position is flawed both empirically and normatively. Empirically, it is just false to claim that *whatever* happens a human being can keep their composure: there are some degrees of suffering that break up the soul from the inside, and make the idea of maintaining one's hardness seem ridiculous.[144] Nor can the capacity for anything like this degree of self-command be developed except under reasonably propitious circumstances.

Normatively, it is at best highly unclear whether the ideal of hardness and self-command is worth pursuing. Nussbaum asks the following:

> What should we think about a human being who insists on caring deeply for nothing that he himself does not control; who refuses to love others in ways that open him to serious risks of pain and loss; who cultivates the hardness of self-command as a bulwark against all the reversals that life can bring? We could say, with Nietzsche, that this is a strong person. But there is clearly another way to see things. There is a strength in the willingness to form attachments that can go wrong and cause deep pain, in the willingness to invest oneself in the world in a way that opens one's whole life up to changes of the world, for good and for bad. There is, in short, a strength in the willingness to be porous rather than totally hard, in the willingness to be a mortal animal living in the world. The Stoic, by contrast, looks like a fearful

---

[144] Simone Weil makes a distinction between mere suffering and affliction (*malheur*) that captures this point well.

person, a person who is determined to seal himself off from risk, even at the cost of loss of love and value. (Nussbaum 1994, p.160)

This criticism obviously applies to Nietzsche as well.[145]

7.4 *Dignity in the Pro-Pity tradition*

But if I am not mistaken, there is another conception of dignity that takes its cue from the pro-pity tradition: let us look at Achilles' reply to the great speech of Priam I quoted above:

> Poor man, you have surely endured many sorrows in your heart. How could you bear to come alone to the ships of the Achaians, into the eyes of a man who has killed many of your brave sons? Your heart must be made of iron. (Book 24, 525 ff.)

Here Achilles responds not just with pity, but with a sort of *awe* – a recognition of *dignity* – at what Priam has done in coming to him.

This feeling is related to pity, but is so to speak its flip side: where pity is a feeling of pain for another like ourselves broken on the wheel of life, the experience of human dignity is a feeling of awe at one like ourselves who has been through adversity, had the vulnerabilities probed, but has *survived without betraying himself.* If pity teaches of the evils that can strike down any human being, dignity uplifts by teaching of the uncanny fortitude that allows people to keep striving for what is worthwhile in the face of destructive blows dealt by life that could have shattered them. We are immensely impressed by such fortitude and think that the human race is ennobled by it.

---

[145] "For a central theme in his work is that Christianity has taught us bad habits of self-insulation and self-protection, alienating us from our love of the world and all of its chanciness, all of its becoming. On this account we have become small in virtue, and will remain small, unless we learn once again to value our own actions as ends, and our own worldly existence as their natural home. I think that in the end Nietzsche fails to go far enough with this critique. He fails, that is, to see what the Stoicism he endorses has in common with the Christianity he criticizes, what "hardness" has in common with otherworldliness: both are forms of self-protection, both express a fear of this world and its contingencies, both are incompatible with the deepest sort of love, whether personal or political." (Nussbaum 1994, p.160)

What Achilles sees in Priam and salutes is very different from what the tradition stemming from the Stoics means by dignity. He is not denying human vulnerability in order to praise strength: rather the sort of strength Priam shows presupposes his own vulnerability and the importance of 'externals' such as securing a decent burial for his son.

On this tradition, it is above all the *hero* who has dignity, and what makes him a hero is precisely his willingness to risk his own destruction for the sake of what is fine.[146] On this way of looking at dignity, it is only to those whom we might feel pity for that we can ascribe a dignity: those who cannot be harmed cannot be dignified in this sense, because for them nothing is at stake. Hence neither Homer's gods (who cannot die or really be harmed) nor the virtuous man of the Stoics (who ensures his invulnerability by curtailing what he finds important)[147] can have real dignity on this conception.

Dignity, on this tradition derives, I suspect, from a reflection on the human condition: our vulnerability, our mortality, our weakness in the face of a world that we cannot control. What impresses us, and makes us think that human life *is* worth affirming after all, is the effort – and the successes – that we see human beings making *in spite* of the adverse and sometimes absurd conditions we find ourselves in. On this way of thinking, the person who has a dignity is above all the hero: and it is in his or her willingness to risk everything for the sake of his or her ideal that this dignity resides. William James encapsulates this tradition well when he writes:

> In heroism, we feel, life's supreme mystery is hidden. We tolerate no one who has no capacity whatever for it in any direction. On the other hand, no matter what a man's frailties otherwise may be, if he be willing to risk death, and still more if he suffer it heroically, in the service he has chosen, the fact consecrates him forever. (1902, p.364)

---

[146] Aristotle's analysis of courage in Nichomachean Ethics III.6-9 is important here.

[147] Epictetus, *Encheiridion* 19: "You can be invincible if you do not enter any contest in which victory is not up to you."

However, as James points out in his discussion, the two main ways in which this will to heroism has been manifested – in war and in religious asceticism – must be non-starters for an account of morality for us now. War is a non-starter because its essential purpose is destruction, and hence whilst it may be a training ground for heroism, its net effect must be for the bad.[148]

Christian asceticism, which James understands as an attempt to cleanse the world through self-mortification,[149] while less destructive than war, is hardly appealing (as James puts it, 'We can no longer sympathize with cruel deities, and the notion that God can take delight in the spectacle of sufferings self-inflicted in his honor is abhorrent.' (1902, p.362)).

James suggests that what we need is a new sort of asceticism that will refocus energies in a healthier way:

> What we now need to discover in the social realm is the moral equivalent of war: something heroic that will speak to men as universally as war does, and yet will be as compatible with their spiritual selves as war has proved itself to be incompatible. (p.367)

His suggestion is that we adopt an ethic of poverty as this moral equivalent of war.[150] Much as I admire William James, I cannot help finding this a bit

---

[148] He quotes an unnamed 'clear-headed' Austrian officer: "Live and let live is no device for an army. Contempt for one's own comrades, for the troops of the enemy, and, above all, fierce contempt for one's own person, are what war demands of everyone. Far better is it for an army to be too savage, too cruel, too barbarous, than to possess too much sentimentality and human reasonableness. If the soldier is to be good for anything as a soldier, he must be exactly the opposite of a reasoning and thinking man. … The recruit brings with him common moral notions, of which he must seek immediately to get rid. For him victory, success, must be *everything*. The most barbaric tendencies in men come to life in war, and for war's uses they are incommensurably good." (1902, pp.366-7)

[149] "It symbolizes, lamely enough, no doubt, but sincerely, the belief that there is an element of real wrongness in this world, which is neither to be ignored nor evaded, but which must be squarely met and overcome by an appeal to the soul's heroic resources, and neutralized and cleansed away by suffering." (1902, p.362)

[150] I cannot resist quoting his claims, which now seem a bit quaint: "We have lost the power even of imagining what the ancient idealization of poverty could have meant: the liberation from material attachments, the unbribed soul, the manlier indifference, the paying our way by what we are or do and not by what we have, the right to fling away our life at any moment irresponsibly, - the more athletic trim, the moral fighting shape. When we of the so-called better classes are scared as men were never scared in history at material ugliness and hardship; when we put off marriage until our houses can be artistic, and quake at the thought of having a child without a bank-account and doomed to manual labour, it is time for thinking me to protest against so unmanly and irresponsible a

silly, for several reasons. First, if we needed a moral equivalent of war, wouldn't the fight against injustice be a better substitute? Second, the view of poverty he is recommending is somewhat romanticised. Third, there is something *gratuitous* about it: one deliberately puts oneself in a situation where one needs to exercise heroism, whereas heroism properly considered is *already called for* by the situation one finds oneself in.

On a more general level, it seems to me that James in addition makes two much more significant mistakes: first, he forgets that the importance of the hero lies not in his ability to justify his own life, but in his ability to inspire awe in others: it is not *the hero* who experiences his own dignity through his struggle, but others who experience the hero's dignity.

Second, it is quite incorrect to think that heroism requires special, contrived circumstances such as voluntarily taken on poverty for its application: everyday life provides sufficient challenges.[151] Ibsen's Dr Stockmann in *Enemy of the People* perhaps provides the literary prototype for the sort of heroism that any honest person may be called upon to pursue in the course of their job.[152] Think of artists and journalists, and of how much courage it can take to portray the truth as they have seen or experienced it unadorned. Or again, think of the often less visible heroism often displayed by people caring for sick or dying relatives. It is true that we 'tolerate no one who has no capacity whatever for [heroism] in any direction', but heroism is always closer to us than we think.

---

state of opinion." (1902, p.368) To be fair to James he later came to the more sensible suggestion of national service as the necessary 'moral equivalent of war'.

[151] I find it a little odd that someone who is usually so finely attuned to the minutiae of normal people's experience should wish to transplant heroism to a completely different realm of life.

[152] More modern cases would include those who risk losing their jobs, or even going to jail, to make public what various branches of the government having been doing in the public's name.

Chapter Eight

Reconstructing Universality

8.1 *Introduction*

This chapter completes the reconstruction of morality by clarifying and extending what we have so far said about dignity and by giving an account of moral rules.

The account of dignity is extended in such a way that we can believe in the dignity of *all* human beings. This takes some work: for the hero is by his or her very nature *exceptional,* so it is immediately problematic to attempt to argue for the dignity of *all* human beings on the basis of the dignity of the hero. We can, as I suggested, extend the category of hero by looking for heroism in more mundane places, but there is a limit as to how far we can do this. For it is implausible to suggest that *every* human being is heroic in some respect: it seems far more realistic to agree, with James that there are some people who have no capacity for heroism in any direction, and some people whose slight heroism in some direction is vastly outweighed by their utter lack of it in all other directions.

My reply to this problem is to make a distinction between two different ways of thinking of the dignity of human beings: a bloodless and a full-blooded, and suggesting that we only need the bloodless version in order to have a viable account of moral rules.

The bloodless conception of the dignity of all human beings amounts to no more than seeing all human beings as worthy of a basic respect: that is considering them to be a member of the cosmopolitan community of morality. Considering them in this way implies three things: i) we consider ourselves to be bound by moral rules in our dealings with them; ii) we

consider them to be bound by moral rules in their dealings with us; iii) we consider them to have a voice in the ongoing discussion as to what moral rules we should adopt for the general regulation of society.

I assume that we can unproblematically hold that everyone has a dignity in this bloodless sense.[153]

The account of moral rules presupposes this bloodless conception of the dignity of human beings. It does three things. First, it explains what the regulation of life through moral rules is supposed to accomplish, namely the continuing viability of morality as a project, and the equal respect of each under rules that all can agree to. Second, it articulates why one might find this goal worth striving towards. Third, it gives a theory of argumentation which we can use to determine which norms to take to be morally valid.

The ease with which we can gain acceptance for this bloodless conception of the dignity of all human beings cuts both ways however. Whilst it allows us to give a meaning to the fundamental moral claims that every human being has a dignity, and that moral rules exist to ensure that this dignity is respected, it fails to do justice to the other motivation we had for insisting on the dignity of all human beings. This was that the dignity of human beings provides the positive affirmation at the heart of morality; the thing that makes it false to claim that morality is basically *reactive* in its system of valuation. Whilst affirming each person as a member of the moral community may go a little way towards fulfilling this goal, I doubt that it is sufficient to perform the affirmation required.

Therefore, I develop a full-blooded conception of the dignity of all human beings alongside this bloodless one. If one were less sensitive than I am to the worry that Nietzsche raises, perhaps one might see it as an optional

---

[153] Note that this account does not rest on any alleged special traits that human beings have; nor do I argue that we *must* hold it. All that it rests on is the legitimacy of adopting this attitude towards them in our dealings with them

extra to the account of morality, and think that the account of moral rules gives us everything we need. But for those who think that morality requires a more ringing endorsement of human beings, it forms an integral part of the reconstruction of morality.

This full-blooded conception of the dignity of all human beings takes its cue from the way we experience the dignity of the hero, but it makes a further suggestion: namely that what is special about the hero – what makes us accord her a dignity – is, more fundamentally than her feats of derring-do, her ability to arrest our usual emotional self-protection in such a way that we cannot help finding her valuable. It follows that if we want to experience the dignity of all human beings in a full-blooded sense, we do better to attempt to dismantle these usual emotional defences than to attempt to make good on the rather implausible claim that everyone is a hero in some respect. I suggest that the most effective way we have of doing this is to first, cultivate a greater openness towards others, and second to adopt an approach to life that places a unilateral love for others at its centre.

The basic structure of the reconstruction of morality, in which moral rules and the affirmation of human life are conceived of as springing from separate sources, and must be combined in creatively in the process of living, is borrowed from Paul Ricoeur. Ricoeur uses this structure to explain what Christianity has to offer ethics. His suggestion is that the Christian ideal of unilateral love should enter into the conception of reciprocity we find in the golden rule and in the universal law formulation of the categorical imperative and transform it by infusing it with a greater generosity: not to *replace* the rule of reciprocity but to *reinterpret* it in a more giving way.[154]

---

[154] "In fact, what penal law and in general what rules of justice could apply directly, *without the detour of the golden rule,* the bare commandment to love one's enemies? What distribution of tasks, roles, advantages, obligations and duties – following the Rawlsian schema of the "idea of justice" – could result from a commandment from which reciprocity appears to have been excluded? What equity, on the economic plane, could be drawn from the commandment to "lend, expecting nothing

What I shall present differs from Ricoeur in two important ways. In terms of Charles Taylor's terminology, Ricoeur's account will place the Christian affirmation of life on the side of the good and the account of moral rules on the side of the right.[155] My account places both the idea of a moral regulation of society and the dignity of human beings on the side of the good, leaving only the actual moral rules to be enforced on the side of the right.

This means that there is far less distance between the two elements on my account.[156] Rather than the gulf we have on Ricoeur's account (which is in addition widened by a further distinction between sacred and profane), we have a distinction between *trying to come to a mutual understanding about which norms can be shared as a basis for cooperation* and *approaching human beings in such a way that one experiences them as having a dignity.* Although the first is a bilateral (or better, multilateral) and the second a unilateral operation, they both come out of a common vision of human life: that of human beings as attempting to create structures wherein they can flourish in a world that is indifferent to their fate.

The second disagreement is already implicit in the first: Ricoeur's account incorporates a much more serious, more 'heavyweight' view of the world

---

in return"? Detached from the golden rule, the commandment to love one's enemies is not ethical but supraethical, as is the whole economy of the gift to which it belongs. If it is not to swerve over to the nonmoral, or even to the immoral, the commandment to love must reinterpret the golden rule and, in so doing, be itself reinterpreted by this rule." (Ricoeur 1991, p.301)

[155] This reading is, I think, true of Ricoeur's *religious* contributions to ethics, but in his secular contributions we have a more subtle three-stage process through which the good and the right are interrelated and reinterpret one another. See Ricoeur (1992, p.170): "According to the working hypothesis I am proposing, morality is held to constitute only a limited, although legitimate and even indispensable, actualisation of the ethical aim, and ethics in this sense would then encompass morality." He states his aim is to establish "(1) the primacy of ethics over morality, (2) the necessity for the ethical aim to pass through the sieve of the norm, and (3) the legitimacy of recourse by the norm to the aim whenever the norm leads to impasses in practice."

[156] It is noticeable that Ricoeur's general philosophical methodology is to approach a problem by setting up two polar opposites and finding a way of mediating between the two of them through the act of creative interpretation. His account of the relationship between Christian ethics and moral rules is no exception. Whilst this strategy is often very revealing and enables him to reconcile seemingly conflicting claims, in this case, he seems to me overplay the conflict, just so he can use his preferred methodology. I prefer to fill out the answer to the question about the affirmation of human life and the answer to the question about the need for universal moral rules in a way that creates much less initial difference between them and so creates less need for creative mediation.

than mine does. His account is anchored in absolutes in both its 'love' and its 'justice' directions: in the love direction, with his reading of Christian teaching,[157] and in the justice direction, with his reading of Alan Donagan's absolutist theory of morality.[158]

Mine is free floating and pragmatist in tone, expressing a faith in the possibility of universal moral rules and in a full-blooded conception of dignity, but well aware that these remain *optional* for us. On my account it is necessary neither to believe in universal moral rules nor in the dignity of all human beings. The underlying value system of morality is, just as Nietzsche said, just one among many other possible systems. But it does not follow either that it is illegitimate to take morality as a guiding light in one's life, or that there is nothing that can be said in favour of so doing.

What speaks in favour of morality is the structure it imposes on human life. I shall suggest that there are three broad things that can be said in favour of morality: i) general adherence to moral rules protects both each individual and the web of interrelations through which they live; ii) belief in universal moral rules allows each person to recognise each person as autonomous and accountable, and thereby creates a domination-free space in which valuable human endeavours are possible. iii) Believing in the dignity of all human beings in a full-blooded sense makes one's encounters with other human beings much richer and more satisfying than they would otherwise be, and hence one finds life more interesting.

Some may find the fact that my account makes morality ultimately optional – that, in Kantian terms it never rises above the level of

[157] Ricoeur bases his account of love on the Christian idea of human beings as *creatures* of God: created by God and utterly dependent on God. For Ricoeur it is from this sense of radical dependence that we find ourselves *summoned* by God and placed in a situation of gratitude and wonder: '"And God saw everything that he had made and behold it was very good." (Genesis 1:31) The hyper ethical dimension of this predicate extended to all creatures is what we must emphasize, for the result is that it is as a creature that we find ourselves summoned. The sense of radical dependence that is at stake here, insofar as it is attached to the symbolism of creation, does not leave us face-to-face with God; rather it situates us within nature considered not as something to exploit but as an object of solicitude, of respect and admiration.' (1991, p. 325).
[158] See Donagan (1977).

hypothetical imperatives – unsatisfactory, perhaps even vertigo inducing. There are two worries: first, whether the account will allow us to deal with those who don't want to be moral; and second whether it is satisfactory for people who *do* believe in morality.

I shall take the second worry first. This worry is about the *contingency* of the account I shall be giving: how can we continue in our commitment to morality after we admit that there is nothing to rationally coerce us into being moral? Can we continue to take morality seriously enough to want to guide our lives according to it?

My reply is twofold: first that it is a sign of moral immaturity to want to be *coerced* – whether by reason or by something else – into being good.[159] Second, there are good reasons why admitting the contingency of morality will make one both *more committed* to fighting the good fight and more effective at so doing than one otherwise would have been.

If one believes that the moral law is inbuilt into every human being, it is easy to become complacent about such matters as moral education: for, try as we might, however evil we attempt to become, we cannot extirpate morality from ourselves. The moral law will still be there to guide each and every human being for as long as the human race survives. But if morality is contingent, then there is much more at stake: there is a very real possibility that the quality of our relations with our fellow human beings could deteriorate to the point where ideas like the dignity of all human beings, or universal rules against murder come to seem like laughable and superstitious naivety.

There is every reason to think that a keen awareness of the contingency of moral goodness will lead one to be more willing to celebrate it and protect it where one finds it, just as a growing awareness of the contingency of

---

[159] In any case, we retain an element of coercion in our account of moral rules – though this coercion comes through *accountability to others* rather than through something more akin to the moral law.

human life has led to a greater desire to protect the ecosystems that make this life possible.[160]

Now, the worry about the amoralist. What are we to *say* to the person who refuses to act morally? My answer is that we do not need an *argument* from neutral premisses that will force them to accept the validity of morality. If *we* do not require such an argument to see the point of being moral, why do we need one for their benefit?[161]

It is sufficient to be able to *educate* such people, so that they can see the point of morality for themselves. If they are ineducable (where, for example, their early experience has damaged them to the point where they have become incurably psychopathic) we may need to find a way of *controlling them* through schemes of reward and punishment that are effective for them.

8.2 *Moral rules*

I shall now proceed to give a pragmatist account of moral rules. On this account, moral rules are social norms which aim to ensure both the viability of the project of morality and equal respect for each under rules that everyone can agree to. Moral rules are held up by no more than a common willingness to believe in interpersonal justice and solidarity: their more precise content is forged by those who are committed to this project, through their attempts to convince one another of what would constitute the set of rules that is best suited to achieve the aims of morality.

Everyone who has any interest in the aims of morality is free to contribute to this debate, by suggesting why individual norms are problematic;

---

[160] As a side issue, I would also argue that recognising the contingency of morality allows us to better understand moral badness, both by allowing us to view it as normal rather than pathological, and by a keener awareness of how one too could have been or could yet become a bad person.

[161] There is the additional pragmatist point that arguments for the validity of morality are, in fact, pretty ineffective at actually changing the minds of bad people: even if one were in possession of a valid argument which showed that someone *really ought* to be moral, the fact is it is unlikely to be effective in changing his behaviour. You will in other words, have to fall back on education or control to get the job done, just as you do on my pragmatist model.

suggesting better ways of formulating them; finding ways of reconciling norms that both seem plausible but which conflict in the behaviour they demand; and in many other ways.

Moral rules presuppose the worthwhileness of a particular ethos, that of living together in harmony with others on the basis of rules that all can agree to, in a community that can eventually be extended to encompass all human beings. Arguments about moral norms presuppose this ethos and only make full sense to those for whom this ethos encapsulates a worthwhile way of living. It is no part of my pragmatist account of moral rules to argue that moral norms can be convincingly justified to anyone for whom this ethos is unimportant.

The account is split into three parts:

1) Explaining what it is that the moral regulation of life attempts to achieve.

2) Attempting to articulate the conception of the good implicit in this ethos.

3) Explaining how, given this ethos, we are to argue about what regulation of life best expresses this ethos.

8.2.1 *What the moral regulation of life aims to achieve*

On any account such as ours, which sees morality as a contingent phenomenon within the history of ethical thinking, the most important task that moral rules must perform is to ensure the continuing viability of morality. The moral rules and the general behavioural expectations that we build into morality must be self-sustaining: moral rules must generate confidence in themselves through people acting on them, and must be such

as to draw increasing numbers of people to hold themselves accountable to them.[162]

There are several reasons why morality might fail to be self-sustaining. The most pressing such reason is that morality might collapse from within. We might find ourselves unable to take morality seriously any more: when, for example, morality tries to get strict with us about our obligations to our fellow human beings, or the importance of truthfulness, we might find ourselves involuntarily smirking at the idiocy of such an idea, much as we would do if someone told us in all seriousness that we should be preparing for the coming of the Last Judgment.[163] I hope to have obviated this potential problem by reconstructing morality in a way that is deliberately post-Nietzschean, and is both modest and sensitive to the structure of human life.

The other threats to morality are external. Put simply, it is much more difficult to act according to moral principles when those around you do not: if you are lucky enough to live in a morally well-ordered community where people generally act according to moral rules and expect others to do the same, then it is relatively easy to act according to moral principles, and vice versa.

We could think of something like a principle of inertia: the general community in which one lives sets a certain tone; and it takes a much greater degree of effort to act against the grain created by this general tone than it does to act in accordance with it.

Moral goodness becomes more difficult to the extent that the tone in a community downgrades moral goodness and expects different responses: in extreme cases, such as under a tyranny of the sort we saw for example in

---

[162] I am in agreement with Annette Baier's claim that 'a decent morality will *not* depend for its stability on forces to which it gives no moral recognition. Its account books should be open to scrutiny, and there should be no unpaid debts, no loans with no prospect of repayment.' (Baier 1985, p.8)

[163] The underlying theme of chapter two was that this has already occurred for much of Kant's account of morality.

Stalin's Russia or in Nazi Germany, the exercise of moral goodness becomes positively dangerous to the good person. Although in such circumstances we still find the odd individual capable of standing up with unbowed courage for the rules of morality, what is notable about such individuals is their very exceptionalness. Morality lives on only on tiny islands in a sea of immorality: in such societies, the project of morality has been shattered.

The first aim of moral rules must therefore be to ensure that such a case does not occur; and, more generally that it is easier, rather than more difficult to exercise moral goodness.

Above and beyond this basic goal of seeking the survival of moral goodness as a way of life, we can think of the purpose of moral rules in a way that is in line with the tradition of morality: namely to ensure that everyone is treated with the respect that is due to them and to impose sanctions on those who violate these requirements of mutual respect.

Where our account differs, however, is that it stipulates that *it is up to us* to decide what we should take to constitute the best way of working out the respect due to each person and how we should ensure that this is met. In 8.2.3 I provide a theory of argumentation which attempts to elucidate what a debate about moral norms should be like if moral norms are to claim our adherence.[164]

---

[164] Habermas has broadly the same conception of the purpose of moral rules, and he suggests, in a passage that is dense and suggestive rather than wholly lucid, that the tie between the project of ensuring the future of morality as a way of life and ensuring of equal respect of each person may be even closer than this: that they *imply* one another. "From the perspective of communication theory there emerges a close connection between concern for the welfare of one's fellow man and interest in the general welfare: the identity of the group is reproduced through intact relationships of mutual recognition. Thus the perspective complementing that of equal treatment of individuals is not benevolence but solidarity. This principle is rooted in the realisation that each person must take responsibility for the other because as consociates all must have an interest in the integrity of their shared life context in the same way. Justice conceived deontologically requires solidarity as its reverse side. It is a question not so much of two moments that supplement each other as two aspects of the same thing. Every autonomous morality has to serve two purposes at one: it brings to bear the inviolability of socialized individuals by requiring equal treatment and thereby equal respect for the dignity of each one; and it protects the intersubjective relationships of mutual recognition requiring solidarity of individual members of a community, in which they have been socialized. … Moral norms cannot protect one without the other: they cannot protect the equal rights and freedoms of the

8.2.2 *Articulating the good implicit in a moral regulation of society*

In order to find the idea of regulation of life in accordance with moral rules attractive, you must believe in two things: first, that human beings are sufficiently valuable to justify ascribing to them a dignity in at least a bloodless sense; and second that that aim of a society regulated according to moral rules is worth achieving.

The bloodless conception of the dignity of human beings is a baseline, beyond which we can think of the dignity of human beings in a more full-blooded way. I shall discuss later how we can attempt to extend this full-blooded understanding of the dignity of human beings so that it covers all human beings: but it is important at this stage to note that moral rules per se, do not depend on this full-blooded reading.

The bloodless conception of the dignity of all human beings, as I have said, amounts to no more than seeing all human beings as members of the cosmopolitan community of morality. Membership of this community implies three things: i) we consider ourselves to be bound by moral rules in our dealings with them; ii) we consider them to be bound by moral rules in their dealings with us; iii) we consider them to have a voice in the ongoing discussion as to what moral rules we should adopt for the general regulation of society.

This bloodless account of the dignity of human beings should be seen as a slimmed down version of Kantian respect. It takes as its key idea *respect for each human being as a rational agent.* Kant explains this in the following terms:

> Every man has a legitimate claim to respect from his fellow men and is *in turn* bound to respect every other. Humanity itself is a dignity; for a man cannot be used merely as a means by any man (either by others or even by himself) but

individual without protecting the welfare of one's fellow man and of the community to which the individuals belong." (Habermas 1990a, p.244) However, given that this claim is not strictly necessary to our project I shall not attempt to pursue it further.

must always be used at the same time as an end. …. But just as he cannot give himself away for any price (this would conflict with his duty of self-esteem), so neither can he act contrary to the equally necessary self-esteem of others, as men, that is he is under obligation to acknowledge, in a practical way the dignity of humanity in every other man. Hence there rests on him a duty regarding the respect that must be shown to every other man. (*MM*, p.462)

This can be conceived as a more abstract form of the recognition we saw in chapter six: but rather than affirming to someone as a concrete individual we are affirming them as a rational, accountable person, and in so doing we are affirmed by them as a rational accountable person. When interacting with someone under the aegis of moral rules, this affirmation of them is funnelled into respect for them as a rational creature who must be treated according to generally valid rules.

On the second point, we presuppose the value of living together under moral rules and a conception of life in which living in harmony with one's fellow human beings is a good in itself. It is difficult to articulate the point of this just because it is so basic. Scanlon has a good try, from the perspective of his theory of morality:

> The contractualist ideal of acting in accord with principles that others (similarly motivated) could not reasonably reject is meant to characterize the relation with others the value and appeal of which underlies our reasons to do what morality requires. This relationship, much less personal than friendship, might be called a relation of mutual recognition. Standing in this relation to others is appealing in itself – worth seeking for its own sake. A moral person will refrain from lying to others, cheating, harming or exploiting them, "because these things are wrong." But for such a person these requirements are not just formal imperatives; they are aspects of the positive value of a way of living with others.
>
> Duty is most familiar in its negative form, in the feeling of unwelcome constraint and the experience of moral guilt. According to the account I am offering, the pain of guilt

involves, at base, a feeling of estrangement, of having violated the requirements of a valuable relation with others. So understood, this familiar negative aspect of morality corresponds to a positive "pull": the positive value of living with others on terms that they could not reasonably reject. (Scanlon 1998, p.162)

Admittedly this value is often occluded, and more often than not it is only when it is threatened with destruction or it turns out that our assumption that we have been living together justly with out fellow creatures turns out to be incorrect that we see how much we presuppose this value.[165]

Hannah Arendt links this idea up in an interesting way with themes from the theory of recognition: for her the most important thing is the communicative space that exists *between* people living together in conditions free from domination. This she calls *power,* and argues that it is the living soul of all political legitimacy:

> Power is what keeps the public realm, the potential space of appearance between acting and speaking men, in existence. … Only where men live so close together that the potentialities of action are always present can power remain with them, and the foundation of cities, which as city-states have remained paradigmatic for all Western political organization, is therefore indeed the most important material prerequisite for power. What keeps people together after the fleeting moment of action has passed (what we today call "organization") and what, at the same time they keep alive through remaining together is power. And whoever, for whatever reasons, isolates himself and does not partake in such being together forfeits power and becomes impotent, no matter how great his strength and how valid his reasons. (Arendt 1958, pp. 200-1)[166]

---

[165] 'Unlike friendship, morality is commonly seen as a form of constraint, not as a source of joy or pleasure in our lives. I am suggesting, however, that when we look at the sense of loss occasioned by charges of injustice and immorality we see it as reflecting our awareness of the importance for us of being "in unity without fellow creatures."' (Scanlon 1998, p.163)

[166] It follows, as I have already suggested, that moral rules must first of all prevent tyranny, as tyranny is the very negation of power. "Tyranny prevents the development of power, not only in a particular segment of the public realm but in its entirety; it generates, in other words, impotence as naturally as other bodies generate political power. This in Montesquieu's interpretation, makes it necessary to assign it a special position in the theory of political bodies: it alone is unable to develop enough power to remain at all in the space of appearance, the public realm; on the contrary, it develops the germs of its own destruction the moment it comes into existence." (1958, pp. 202-3)

It is important to see all these moves for what they are: partial attempts to elucidate what lies at the centre of the project of morality – what Kant called the kingdom of ends.[167] But rather than adopting a strictly Kantian interpretation of this, we have filtered it through our different understanding of human life.

8.2.3 *How we should decide on what moral rules are valid*

On our pragmatist model, the only viable option to pursue in attempting to work out what the precise content of the rules we should adopt to ensure the continuance of morality as a way of life and the equal respect of all, is to rely on an ethics of argumentation (or, *discourse ethics*), that i) attempts to lay out the rules of a fair discussion as to the validity of moral norms and ii) leaves the actual validation of moral norms to the results of such substantive discussions.[168] For there is nothing left to base the validity of moral norms on other than the agreement of those who are to have their conduct regulated by these norms.

Translating this thought into the moral realm, Habermas argues that our first principle of validation should be (D): 'only those norms can claim to be valid that meet (or could meet) with the approval of all affected in their capacity *as participants in a practical discourse.*'

Discourse is an idealized conversation oriented towards reaching mutual understanding. Robert Alexy provides the fullest working out of what such

---

[167] cf *G* pp.433-4: "For all rational beings stand under the laws that each of them should treat himself and all others never merely as a means but always at the same time as an end in himself. Hereby arises a systematic union of rational beings through common objective laws, i.e., a kingdom that may be called a kingdom of ends (certainly only an ideal), inasmuch as these laws have in view the very relation of such beings to one another as ends and means. ... Hence morality consists in the relation of all action to that legislation whereby alone a kingdom of ends is possible."

[168] Such an approach is in a sense deeply still Kantian, but it is rather the Kant of the *Critique of Pure Reason* and of *What is Enlightenment* that is being appealed to. Cf (*C1*, A738-9/B766-7): 'Reason must in all its undertakings subject itself to criticism; should it limit freedom of criticism by any prohibitions, it must harm itself, drawing upon itself a damaging suspicion. Nothing is so important through its usefulness, nothing so sacred, that it may be exempted from this searching examination, which knows no respect for persons. Reason depends on this freedom for its very existence. For reason has no dictatorial authority; its verdict is always simply the agreement of free citizens, of whom each one must be permitted to express, without let or hindrance, his objections or even his veto.' See also Onora O'Neill (1989, chapters 1 and 2).

a discourse requires: I shall follow his codification of the rules of discourse here.

The first set of rules ensure the logical consistency of arguments about norms: they are as follows.

> (1.1) No speaker may contradict him or herself.
>
> (1.2) Each speaker may only assert what he himself believes.
>
> (1.3) Each speaker who applies a predicate F to an object a, must also be prepared to apply F to any other object which is similar to a in all relevant respects.[169]
>
> (1.4) Different speakers may not use the same expression with different meanings. (Alexy 1990, p. 163)

The purpose and validity of these restrictions should be self-evident: (1.1) follows from the basic rules of logic. (1.2) Is held to be constitutive of the process of serious argumentation. It does not 'exclude the utterance of conjectures, it requires merely that they can be characterised as such.' (ibid, p. 164) (1.3) Demands that each speaker must use terms consistently; (1.4) is designed to ensure that argument is not hampered by lack of common definition of key terms.[170]

Next, there are requirements concerning the openness of the discourse:

> (2) Every speaker must justify what he or she asserts upon request, unless she can provide grounds which justify avoiding giving a justification.
>
> (2.1) Anyone who can speak may take part in a discourse.
>
> (2.2) (a) Anyone may render any assertion problematic.
>
> (b) Anyone may introduce any assertion into the discourse.

---

[169] As applied evaluative claims, this becomes (1.3′) "Any speaker may only assert such value and obligational judgments as he would equally assert in all situations which are the same in all relevant respects to the situation in which he or she makes the assertion."

[170] Doing this may prove problematic and require conceptual analysis or other clarificatory procedures. The legitimacy of such procedures is encoded at (4.2).

(c) Anyone may express his/her opinions, wishes and needs.

(2.3) No speaker may be prevented by constraint within or outside discourse from making use of his/her rights established in (2.1) and (2.2). (Alexy 1990, p. 166-7)

The purpose of these rules is to ensure that any consensus that we come to as a result of such a process of discourse can fairly be said to be binding on all the participants. However these rules are problematic inasmuch as 'it is precluded on factual grounds that all persons utilise their rights standardised in (2.1) and (2.2) and it can be questioned whether the absence of constraint demanded by (2.3) can ever be achieved' (Alexy 1990, p.167)

Therefore, we must see rules as *idealisations* that inform practice, and which we attempt to make our actual practice conform to as closely as is feasible; not as things that can actually be achieved in an actual process of argumentation. It follows that we need to make some rules of the burden of argumentation which alleviate this problem in a reasonable way by making presumptions about what needs to be argued for and what does not.

There are four central ideas here: first that we combine (1.3′) with (2) in such a way that the burden on argumentation must rest with anyone who wishes to argue for *unequal* treatment:

(3.1) Whoever wishes to treat a person A differently from a person B is obliged to justify this.

Second, we assume, via what Perelman called the principle of inertia, that 'an interpretation of practice that is once accepted may not be given up again without reason.'[171] It follows from this that if a statement or norm is presumed to be valid within a particular community, it can only be doubted if a reason for doubting it can be provided.

Third, although (2) allows any speaker to 'problematize any assertion without restrictions', we make the pragmatic rule that once a speaker has

---

[171] Alexy (1990, pp. 167-8); referring to Perelman and Olbrechts-Tyteca (1970, p. 142).

206

given argument in favour of a proposition, someone cannot ask for further justifications of that principle unless she has a couter-argument to show that this argument is deficient. Alexy formalizes these points as follows:

> (3.2) Whoever attacks a statement or norm that is not the subject of discussion must provide a reason for doing so.

> (3.3) Whoever has put forward an argument is only committed to further arguments in the case of a counterargument.

Fourth, and finally, the rules under 2 allow any speaker to filibuster, distract the conversation down irrelevant parts and so on. Whilst this need not be a problem if it occurs occasionally, if it occurs often in could derail discourse. For this reason Alexy suggests

> (3.4) Whoever introduces an assertion of a statement concerning his opinions, wishes, or needs into the discourse, which as argument is not related to a previous statement, has to justify upon request why he/she has introduced this assertion of this statement.

We require further rules, by means of which we can switch to a more self-conscious discourse in order to iron out any problems that might arise at a first level approach: these problems are first of all, disagreements over factual premisses; difficulties with different speakers using terms in different ways or in unclarities surrounding key terms; and last disputes about the very structure of discourse itself. This gives rise to the following rules:

> (4.1) It is possible at all times for any speaker to switch to a theoretical (empirical) discourse.

> (4.2) It is possible at all times for any speaker to move to a linguistic-analytical discourse.

> (4.3) It is possible at all times for any speaker to move to a discourse on discourse theory. (Alexy 1990, pp. 175-6.)[172]

---

[172] I have renumbered these; they appear as 6.1, 6.2 and 6.3 in his account.

Between them, these rules attempt to establish a framework within which we must proceed if we are to take the results of a discussion to be generally binding. It aims to be a theory of argumentation as such: but what makes a moral norm discursively redeemable is a bit more complex. Obviously valid moral norms must meet Habermas' (D): 'only those norms can claim to be valid that meet (or could meet) with the approval of all affected in their capacity *as participants in a practical discourse.*' But in and of itself this lacks sufficient content to delimit what we are attempting to achieve in arguing about moral norms.

I have two modest suggestions as to how we should proceed: first, we should adopt Habermas' thesis about when a norm is morally valid:

> A norm is morally valid when the foreseeable consequences and side effects of its general observance for the interests and value orientations of *each individual* could be freely accepted *jointly* by *all* concerned'.[173]

Second, we keep in mind that our task does not require us to create a completely new system of moral norms *ex nihilo*: the tradition of morality provides us with two very important resources. There are a large number of precepts that have been worked over, argued for and modified in the course of the tradition, for example, those concerning the conditions under which it is permissible to lie. Following the principle of inertia, we can take such precepts as valid until such time as a convincing objection or counterargument is brought to them. Further, the tradition provides us with a vast storehouse of ideas and analogies which we can helpfully redeploy in our current discourse about moral rules.[174]

---

[173] However, Scanlon's contractualist principle, which states that an act is morally permissible if it would be "permitted by principles that could not reasonably be rejected, by people who were moved to find principles for the general regulation of behaviour that others, similarly motivated, could not reasonably reject" (1998, p. 4) is also a plausible contender.

[174] A good starting point for what the tradition of morality can bring to discourse ethics is Donagan's *The Theory of Morality* (1977). Donagan is highly sensitive to the tradition of morality, writes within it consciously and puts forward an account of moral rules which aims to synthesize the results of this tradition.

It is not my place here to usurp the role of such a discourse. It is sufficient for my purposes to have shown how, even on this pragmatist account of morality, we can still see the attractiveness of a society regulated by moral rules, and still have sufficient resources to explain how we can think of such rules as valid.

8.3 *Believing in the dignity of all human beings in a full-blooded sense*

I shall now attempt to explain how we can extend the full-blooded conception of dignity so that it covers all human beings. The initial problem in so doing is, as I have already indicated, that the story we put forward in the previous chapter seems to account for the dignity of only the hero; and it seems difficult to imagine how we could extend it to cover *all* human beings, for it is highly implausible to suggest that *all* human beings are heroic.

This problem is based on a misunderstanding: the figure of the hero was introduced because he or she provides the most obvious example of the way in which we can and do find flesh and blood human beings valuable; that is, capable of justifying human life. I did not mean to suggest that *the only* way that we can experience other human beings as having this sort of value is through seeing them as heroes in some respect or other.

I shall briefly consider the idea of the hero, and then go on to suggest how we should rethink the attempt to believe in the dignity of all human beings.

What is special about the hero is that we *identify* with her: we take her struggle and her courage as emblematic for our own struggle, and are inspired by it. In normal life, we have various emotional defences against identifying with people in this way: but the hero manages to disarm these emotional defences to being impressed by another.

I think that further reflection shows that there are two ways in which we can understand our relationship to the hero: a false way and a true way. If

we understand the hero falsely, we use the hero's story as a way of confirming ourselves in our own worth: we identify with her in her struggle for existence only to the extent that we can assimilate what she has done into the pattern of our own life. In being impressed by her we are doing no more than to be impressed at our own possibilities: seeing the dignity of the hero can be no more than a veiled form of self-glorification or bravado.[175]

The true way of appreciating the hero never lets its awe at the hero cloud its awareness that the hero is a different, concrete individual. It is aware of the varieties within heroism and the way life offers different opportunities for heroism. It leads not toward self-glorification but towards an appreciation of different people in their concrete individuality. This conception of heroism leads towards a belief in the dignity of all human beings rather than away from it. By bringing into play the variety of heroes, it nudges us in the direction of a greater curiosity about human beings and also a greater love for them.[176] It is ultimately this curiosity and this love than we need to cultivate within ourselves if we are to believe in the dignity of all human beings.

Suppose we back up a little now and ask a question about those whom we do *not* see as justifying life through their very presence: what is it that stops us from finding them valuable? The obvious answer is that we are seeing such people aright as the mediocre and uninspiring people they are, and that there is nothing more to be explained.

But this answer seem obvious only from a certain perspective. For there is another alternative that we find in the tradition of morality: namely the

---

[175] This is perhaps most obviously the case where violent and inadequate people identify with violent screen characters such as Rambo. Brian Keenan has an interesting discussion of the way his captors in Lebanon identified with Rambo and with the characters in the A-team. (Keenan 1992).

[176] On this conception, what we need is not more Rambos or James Bonds, but more great listeners who are able to divine the contours of other people's ordinary and individual lives and communicate them to us. In this regard, I would take the work of Studs Terkel (e.g. 1972, 1992) and Tony Parker (e.g. 1983, 1990) to be exemplary.

suggestion that it is the person's self-absorption – inability to look beyond his own nose – that makes others seem uninspiring. On this view, what one finds interesting and inspiring is closely related to what one focuses one's attention towards. (Love is considered to be the highest form of attention). If one focuses one's attention exclusively on one's self and one's own concerns, then it is only natural that one does not find the concerns and lives of others interesting or inspiring. However, if one spends less time focusing on oneself and more on others then one effortlessly finds them interesting and inspiring. On this view, the fact that one finds the majority of human beings dull and uninspiring is an indictment of you, not of those you find dull: it is your self-absorption rather than their mediocrity that is the problem.

When this claim has been put forward in the past, it has usually been done within a realist or a Platonist framework, which ends up begging the question against those who find their fellow humans dull. I suggest we situate it instead within the framework of Williams James' ethics of belief. I shall suggest that James shows us that we have every right to adopt this outward-focused approach to life, and that if we do so, we can see the hero as exceptional in a different way: not as superhuman and making the rest of humanity seem inadequate in comparison, but rather someone whose value as a human being is just particularly difficult to ignore, and which penetrates even our normal willed blindness.

8.3.1 *The right to believe in the dignity of all human beings*

In 'The Will to Believe', Williams James argues that there are certain sorts of proposition that it is legitimate to believe, and to act on the basis of believing, even though one has not been intellectually coerced into so believing. There are three classes of proposition to which he wishes to apply this right to believe.

Most simply, there are cases where we have to act without a full knowledge of the facts, because waiting to ascertain the full facts would be to miss the moment at which action was appropriate or would effectively involve a decision *not* to act; a decision that was just as unjustified as the one to act. Imagine for example I'm on a crowded tube train and I think I see a man removing the wallet from another man's pocket. *I think* I saw him do it; but I'm not sure: was he just pushed into the other man by a jolt of the train?

If I wait to see if there is any further evidence to confirm or disconfirm my thought, it's odds on that one or the other man will get off, and I will have missed the moment for action: by waiting to see, I will in effect have decided *not* to challenge the would-be thief. Of course, if I speak up, I risk accusing an innocent man, offending him, perhaps even getting into a fight with him if he stands on his honour. I take it that situations involving this sort of uncertainty are relatively common in everyday life.

James suggests that situations like this show us that there is a difference between two possible ethics of belief: one which has as its fundamental law *we must know the truth,* and the other, which has as its law *we must avoid error*: 'We may regard the chase for truth as paramount, and the avoidance of error as secondary; or we may, on the other hand, treat avoidance of error as more imperative, and let truth take its chance.' (1897, p.209) In the case we are considering, the truth favouring option would lead us to challenge the would-be thief, believing it worth the risk of being wrong: whilst the error avoiding option would lead us to not say anything unless we were absolutely sure.

James argues against the error avoiding option both because it over-estimates the danger of believing and acting on falsehoods and in addition impedes the search for truth:

Our errors are surely not such awfully solemn things. In a world where we are so certain to incur them in spite of all our caution, a certain lightness of heart seems healthier than this excessive nervousness on their behalf. (1897, p.210)[177]

A second class of case where it is legitimate to exercise the will to believe is where we must take up a stance, but where no evidence we could have in our possession seems at all likely to coerce the intellect one way or the other. For James, the question of the attitude we are to adopt towards religion is of this sort.[178] It is legitimate to choose to have religious faith at one's own risk because *both* the opposing alternatives – agnosticism and atheism – are equally as unable to coerce the intellect, and choosing either will have an attending chance of losing the truth.[179] Given this uncertain situation it is legitimate for each person to choose for himself, according to what he feels most drawn to, so long as he chooses at his own risk.

Last, and most important for our purposes, there are cases where the will to believe applies in a more radical way. There are cases where what we believe and the sheer fact of our acting on it is not so much a matter of plumping for an interpretation under conditions of uncertainty, but partly *makes* the facts in question. Many situations involving the interrelations of human beings are of this kind. For example, whether you like me or not is not independent of the attitude I take up towards you in my dealings with you: if (like the error avoider) I await some sort of proof that you like me before I am willing to countenance being friendly towards you, we are far

---

[177] There is an ethical reason lurking only just below the surface: the truth favouring option is the hero's choice; the error favouring one the coward's choice.

[178] James gets to this conclusion by defining the crux of religion in the following rather thin way: the religious hypothesis says two things: (1) 'the best things are the eternal things, the overlapping things, the things in the universe that throw the last stone, so to speak, and say the final word.' (2) 'we are better off even now' if we believe the first claim to be true. (1897, pp. 214-5) Given this slimmed down conception, it is at least plausible to imagine that the first claim 'cannot yet be verified scientifically at all.' (1897, p. 215)

[179] "Our passional nature not only lawfully may, but must, decide an option between propositions, whenever it is a genuine option that cannot by its nature be decided on intellectual grounds; for to say, under such circumstances, 'Do not decide, but leave the question open,' is itself a passional decision, - just like deciding yes or no, - and is attended with the same risk of losing the truth." (1897, p. 205)

less likely to end up friends than if I start by acting on the faith that you will like me and will return my friendliness.[180]

The general idea is that 'the desire for a certain kind of truth here brings about that special truth's existence; and so it is in innumerable cases of other sorts. Who gains promotions, boons, appointments, but the man in whose life they are seen to play the part of live hypotheses, who discounts them, sacrifices other things for their sake before they have come, and takes risks for them in advance? His faith acts on the powers above his as a claim, and creates its own verification.' (1897, p.213)

Belief in the dignity of human beings in a full-blooded sense is legitimate in both the last two senses of the will to believe: i) faith in the dignity of all human beings is not ruled out by any facts we might discover about the world, and ii) faith in the dignity of all human beings helps to create the conditions for its own verification: through presupposing that everyone is valuable and worthy of attention we most often find our presuppositions confirmed.

Faith in the dignity of all human beings in the full-blooded does not rest on any alleged metaphysical facts – such as humans having an immortal soul or being made in the image of God. Nor does it invoke any special trait that might turn out not to be shared fully amongst humans, or to be shared too widely. All it relies on is the ability of humans to impress and inspire one another: this they can do through several means, most obviously through heroism. It does not strike me as plausible that any further facts discovered could show that we have no *right* to be inspired in this way by other human beings, any more than any further facts could show that we have no right to feel pity or anger.[181]

---

[180] "Whether you [like me or not] depends, in countless instances, on whether I meet you half-way, am willing to assume you must like me, and show you trust and expectation. The previous faith on my part in your liking's existence is in such cases what makes your liking come. But if I stand aloof, and refuse to budge an inch until I have objective evidence… ten to one your liking never comes. (1897, p.213)

[181] As I argued in the previous chapter, the dignity of human beings can still be a source of moral

Faith in the dignity of human beings is like trust: by believing it and spreading this belief through the way we act, we alter the terms of engagement on which we meet our fellow human beings. Human beings living in an environment where faith in the dignity of human beings is powerful tend to recreate the conditions of possibility for that faith by both acting in a way that is expressive of their own dignity and by respecting the dignity of other human beings. Conversely, human beings living in an environment where faith in the dignity of human beings is weak or nonexistent tend to recreate the conditions that make faith in the dignity of human beings difficult by both acting in a way that is undignified (mean spirited etc) and by not respecting the dignity of others.

We have every right to believe in the dignity of human beings, then, for two reasons: first because there are no facts that stop us from doing so, and second, because a preliminary faith in the dignity of human beings is necessary if we are to create societies in which the dignity of human beings can readily be experienced. This is a case where 'faith in a fact can help create the fact' (James, 1897, p.214): certainly we cannot stand aloof from human beings, take little interest in others and *wait to see* if others *really are interesting*. One must take the plunge.[182]

So it is a legitimate project to have a prior faith in the dignity of all human beings in this full-blooded sense.

But experiencing the dignity of another human being involves openness: a willingness to bracket one's preconceptions, what one thought one knew for sure, and risk real genuine contact. If we are insecure; if we fear being

uplift and moral motivation, but by explaining the workings of dignity in an *existential* sense (that is, relative to the person's understanding of the vulnerability of human life, and the other person's courage in carrying on living in the face of this vulnerability) we obviate the need for any account of what is sacred in a human being, or of the special trait that gives human beings their dignity.

[182] 'We have a story in India about two men, one high-minded and generous, the other very selfish, who were sent to foreign lands and asked to tell what kind of people they found there. The first reported that he found people basically good at heart, not very different from those at home. The second man felt envious hearing this, for in the place he visited everyone was selfish, scheming and cruel. Both, of course, were describing the same land.' Easwaran 1986, p. 66.

exposed as fool's gold by the touchstone of experience, then we will try our hardest to avoid situations where we must respond spontaneously to another human being. If we are afraid that really meeting another will make us doubt something that we thought we knew for sure, then we will avoid really meeting them. If we feel that we need to hold tight to ourselves to stop ourselves spinning off centrifugally into a thousand pieces, then we will avoid encountering another as possessing a dignity.[183]

Following this path can be painful and difficult so few try it; but if we pursue this path, then we can come to appreciate that much of our lives are usually spent in a willed blindness to the value and experiences of others.

What marks out both curiosity about and love for human beings is that they involve a focusing our attention *outwards* – away from our selves and our own concerns and onto other human beings. They pour interest and appreciation into the world: but their effort is not wasted, but rather pays great dividends by making life much richer and more interesting. Where there was fear and self-loathing, they create a sense of freshness and possibility.

The hero comes to seem exceptional in a different way: not as superhuman and making the rest of humanity seem inadequate in comparison, but rather someone whose value as a human being is just particularly difficult to ignore, and which penetrates even our normal willed blindness.

---

[183] Cf Velleman "[Love] arrests our tendencies toward emotional self-protection from another person, tendencies to draw ourselves in and close ourselves off from being affected by him. Love disarms our emotional defences; it makes us vulnerable to the other. … Many of our defences against being emotionally affected by another person are ways of not seeing what is most affecting about him. This contrived blindness to the other person is among the defences that are lifted by love, with the result that we really look at him, perhaps for the first time, and respond emotionally in a way that's indicative of having really seen him." (1999, p.361)

8.3.2 *How to overcome our defences through the cultivation of love*

Following this path leads to an ethic which takes as its key idea unilateral love expressing itself in spontaneous generosity, which incorporates a present-oriented conception to time, in which one attempts to live as fully and as generously as possible, holding nothing back. I consider in 8.4 how we should relate this conception of life back to that which finds expression in the attempt to construct moral rules.

The account of love that we need is one that reworks the Christian ideal of love as exemplified in the Sermon on the Mount. The two key features about this conception of love are, first and more familiarly, its unilateral nature: its willingness to make the first move and to love without needing this love to be returned.[184]

Second, its present-oriented conception of time, that banishes both anxiety about the future and recrimination about the past so that the present becomes an open space of possibility. Anxiety about the future is banished by placing one's life in God's hands: [185]

> Rather than something futural, this is a *presential* time, a time of presencing, which lets today be today. Trusting oneself to God's rule, the day is not drained of its time. Today is not sacrificed to tomorrow, spent in making oneself safe and secure against tomorrow. It is a temporality of trust, of trusting oneself to God's rule, and in so doing to time and the day. (Caputo 1999, p.99)

---

[184] See for example Matthew 5.43-8: "You have heard that it was said, 'You shall love you neighbour and hate you enemy,' But I say to you, Love your enemies and pray for those who persecute you… For if you love those who love you what reward have you? Do not even tax collectors do the same? And if you salute only your bretheren, what more are you doing than others? Do not even the Gentiles do the same? You, therefore, must be perfect, as your heavenly Father is perfect."

[185] Cf Matthew 6.25-27: "Therefore I tell you, do not be anxious about your life, what you shall eat or what you shall drink, nor about your body, what you put on. Is not life more than food, and the body more than clothing? Look at the birds of the air: they neither sow nor reap nor gather into barns, and yet your heavenly Father feeds them. Are you not of more value than they? And which of you by being anxious can add one cubit to his span of life?"

And we are loosened from the grip of the past by the power of forgiveness: "Forgiveness keeps the net of social relationships open and makes possible what Arendt calls "natality," the fresh, natal, initiating power of a new action, new beginnings, new starts. Each day is a new day, a renewal of the day, a new gift." (Caputo 1999, p.96-7).

> By trusting God's rule one breaks the chain of time and frees the day, letting the day come-to-presence, tearing up the chain of time, freeing it from the circulation of debts and anxieties, letting the day be a "gift". Forget what is owed to you in the past; forget what you owe to the future; tear up the chain of time and take today as a gift… It is, something new and freeing has begun now which is now with us and frees us from the past and future. (Caputo 1999, p. 101)

When we put these two elements – the unilateral love and the presential time – together, we get a way of living in which all the usual chains that bind us; via time and via other's expectations of us are smashed, and all that is left is an outflowing of generosity and joy.

I am not suggesting a return to a Christianity – even to a mystical, half-deconstructed form of it however. What I am suggesting is a *reappropriation* of this ethic within a secular context.

Rather than being grounded in the gift of God, love becomes utterly groundless, gratuitous, and is embraced in its very gratuity. Cupitt expresses this thought in a way that I am happy to agree with:

> Life cannot be possessed or clutched at: we should pour ourselves out and pass on, without hesitation or regret. We can get ourselves together only by leaving ourselves behind. That is solarity – to live by dying all the time, heedless, like the sun and in the spirit of the Sermon on the Mount. Solar ethics is a radically emotivist and expressionistic reading of the ethics of Jesus (Cupitt 1997, p.90)

Is such a love possible without an underlying Christian faith? For, on a classical Christian account this love wells up from an awareness of our status as God's creatures and our dependence on God.

I myself see no objection to treating belief in God as a ladder that we can push away once we have attained to the moral understanding implicit in this ethic. We could approach this thought from two directions: either from within Christianity, as a radicalisation of the Christian message in a direction first suggested by Christian mysticism and especially Meister Eckhart, and most interestingly worked out in a series of works since 1980 by Don Cupitt[186] or from the direction of atheism, which would opine that, given that i) there is no God, and ii) people have in fact been able to act in accordance with the demands of a unilateral love nonetheless, then iii) the non-existence of God is therefore no barrier to the possibility of unilateral love; and iv) one can therefore practice this even without a belief in God.

8.4 *Moral life*

The account of moral rules and the full-blooded account of dignity should, I hope, complement rather than undermine one another. We have on the one hand, the project of trying to come to a mutual understanding about which rules we should use to regulate our life together; on the other, the project of attempting to restructure one's way of interacting with human beings through a reworking of the Christian ideal of love.

---

[186] Cupitt publishes at least one book a year, and in addition is always moving in his thought, so it is a little difficult to recommend any one book as representative of his oeuvre. What we have is a journey, from a fairly orthodox Church of England position, to doubts becoming ever more radial, which he understood in a Kierkegaardian way, to a final break with any sort of realist account of Christianity (1980). Then in the mid-eighties, an attempt to construct an anti-realist version of Christianity, to a break with even anti-realist Christianity towards a flirtation with an anti-realist reading of Buddhism, to a point where he now advocates utilising certain elements from Christianity and Buddhism in the service of interesting experiments in selfhood: "I am suggesting that we can and should now be uninhibited and eclectic in creating new religious meanings, practices, and narratives out of the materials available to us. The poetical theology will fiction and refection our religion, tell and retell the old stories. What will make it a *theology* will be its use in helping us to see ourselves and our life with a greater clarity of moral vision, in helping us to be "easy, going" about the transience of everything, and in showing us how to live ardently." (1997, p.120)

The first is a rule based ethics of mutual accountability; the second an overflowing ethics of generosity. This 'solar ethics' goes beyond anything that might be required on the basis of moral rules: it goes beyond rules, beyond mutual obligation into the ecstatic space of unilateral love. This unilateral nature of it is its greatness: its gratuitous kindness and appreciation makes it able to make the first move, to create the conditions for cooperation and mutual understanding out of hostility or indifference. But its weakness is its lack of consistency.

The strength of moral rules is precisely their ability to bind people together into a community of understanding through which tasks can be apportioned, offenders punished and so on. Their weakness is their tendency to efface the importance of each individual and the differences between them and to slump down to a merely perfunctory level at which the universally shared expectations become little more than a figleaf for indifference to others.

A life of true moral goodness requires both: from the creative tension between them we create the true life. It would not, I suspect, be a life that Nietzsche would have been happy to have lived, but yet it embodies a convincing reply to his claim that life needs to be justified, doing so both through the ardency of its solar living and through the structures of mutual recognition created through the moral rules it upholds.

AFTERWORD

The major difference between my account of morality and all the others we considered is that my account does not attempt to 'ground' morality – that is, attempt to show that there is an argumentative route from premisses no one could reasonably doubt to the validity of morality. Rather it is geared to showing the *legitimacy* of confidence in morality and in finding ways that build this confidence in a realistic way.

Whilst the three main pillars of morality – cosmopolitanism, reciprocity and the dignity of human beings are all still very much standing, morality as I have reconstructed it is much *lighter* than on the other accounts: there is no transcendent source of morality; no a priorism; no irremisible voice of duty. Rather we have a system that is human and fallibilistic all the way through; a system that reconstructs morality out of elements of human beings as we know and interact with them: our vulnerability; our potentiality for faith and our willingness to reach a mutual understanding.

How we interpret this change depends on how we view the project of grounding morality: if we think that that project still makes sense, then one is likely to think that the structure I have erected is at best a temporary stop-gap, a way of showing that even though we do not currently have a sound argument demonstrating the necessity of morality, there are still some quite significant things that can be said in favour of the way of life of believing in morality.

If, however, one thinks that the project of grounding morality has become intellectually bankrupt, then the structure I have erected will seem to be the only honest way of proceeding: to be the best way of ensuring a future for morality.

I have come over time to alter my position from the first interpretation to the second. I do not know what to say to make this interpretation plausible to others, or even how to fully explain my change of mind. It has a lot to do with the status one accords contingency in human life. Much of my earlier thought was connected to the question of how to overcome the contingency of human life – how to show that despite the fact that we human beings are the contingent result of a long process of natural selection, and despite the fact that the ideas that make up the core of morality have a history and undergo various transformations in this history, every human being *ought to* accept the moral regulation of life, because this is required by practical reason.

The most obvious way to do this would be to construct an account at the level of the right.[187] And so I spent some time trying to construct an account of practical reason that would allow us to demonstrate that *everyone ought to be moral, because moral standards really are normative.* But a combination of the work of Nietzsche and Taylor undermined this hope in me, for reasons that I have set out in chapters three and five: no account at the level of the right can explain why we should take these purported moral truths seriously, and it is this question which is prior to that of the right.

However, if we are working at the level of the good, as I hope the chapter on Taylor makes clear, it's difficult to see how we could ground morality in default of an ultimate belief in God or in some other metaphysical principle.

I thus found myself in an aporia: I literally didn't know which way to turn. So I backed up, and I began to unpick what it was that led me to want to achieve this certainty, this ability to demonstrate that everyone must be

---

[187] One can bolster one's confidence in a way of overcoming contingency by the analogy with mathematical truths: perhaps, just as natural selection should not undermine our belief in the objectivity of mathematics, so it shouldn't undermine our belief in the objectivity of moral truths.

moral. In this I found Rorty's thought helpful above all. Rorty suggests that what lies behind the attempt to ground morality is, as Nietzsche says, *resentment* at the more powerful getting away with committing outrages against the weaker:

> The residual popularity of Kantian ideas of "unconditional moral obligation" – obligation imposed by deep ahistorical forces – seems to me almost entirely due to our abhorrence of the idea that the people on top hold the future in their hands, that everything depends on them, that there is nothing more powerful to which we can appeal against them. … We *resent* the idea that we shall have to wait for the strong to turn their piggy little eyes to the suffering of the weak, slowly open their dried up little hearts. (1993, p.182)

What we are really attempting to do in trying to ground morality is to construct a larger, more powerful stick with which to beat them:

> We desperately hope there is something stronger and more powerful that will *hurt* the strong if they do *not* do these things – if not a vengeful God, then a vengeful aroused proletariat or, at least, a vengeful superego or, at the very least, the offended majesty of Kant's tribunal of pure practical reason. (1993, p.182)

If this is so, then "the confusion of ideals and power, is *all* that lies behind the Kantian claim that it is not only nicer, but more *rational,* to include strangers within our moral community than to exclude them." (1993 p.182) At first I found this suggestion rather disquieting and irritating (as a piece of typical Rortian flippancy), but I have come over time to accept it as an interpretation of what led me towards the attempt to justify morality in the first place.

This has freed me up to approach morality, and the arguments we give in favour of it in a different way: one that I hope it will gain in effectiveness from its greater honesty and self-awareness as to its own motives.

I started this project with a deep worry, namely how to reconcile my will to believe in morality with my will to truth. I can report that I have at least succeeded in laying to rest my own worries: whether in so doing I have betrayed either morality or the will to truth I leave the reader to decide.

BIBLIOGRAPHY

Alexy, Robert (1990). 'A Theory of Practical Discourse', in Benhabib and Dallmayr (1990).

Altham, J. E. J, and Harrison, Ross eds. (1995). *World, Mind and Ethics: Essays on the Ethical Philosophy of Bernard Williams.* Cambridge: Cambridge University Press.

Äpel, Karl-Otto (1997). 'Kant's "Toward Perpetual Peace" as Historical Prognosis from the Point of View of Moral Duty' in Bohman and Lutz-Bachmann (1997).

Arendt, Hannah (1958). *The Human Condition.* London: University of Chicago Press.

Aristotle, *Nichomachean Ethics* trans. Terence Irwin (1985). Indianapolis: Hackett Publishing Company.

—— *Metaphysics,* in *The Complete Works of Aristotle,* ed. Jonathan Barnes. Oxford: Oxford University Press (1984).

—— *Rhetoric,* in *The Complete Works of Aristotle,* ed. Jonathan Barnes. Oxford: Oxford University Press (1984).

Baier Annette C. (1985). 'What Do Women Want in a Moral Theory?', *Nous* vol. 19, 1985 pp. 55-63, reprinted in Baier (1994).

—— (1994). *Moral Prejudices: Essays on Ethics.* USA: Harvard University Press.

Baynes, Kenneth, Bohman, James and McCarthy, Thomas eds. (1987). *After Philosophy: End or Transformation?* Cambridge, Mass.: MIT Press.

Benhabib, Seyla and Dallmayr, Fred eds. (1990). *The Communicative Ethics Controversy.* Cambridge, Mass.: MIT Press.

Benjamin, Jessica (1988). *The bonds of love: Psychoanalysis, feminism, and the problem of domination.* New York: Pantheon.

—— (1995). *Like subjects, love objects: essays on recognition and sexual difference.* New Haven: Yale University Press.

Bohman, James and Lutz-Bachmann, Matthias eds. (1997). *Perpetual Peace: Essays on Kant's Cosmopolitan Ideal.* Cambridge, Mass.: MIT Press.

Cicero, *De Republica.* Translated by Niall Rudd in Rudd (1998).

—— *De Legibus.* Translated by Niall Rudd in Rudd (1998).

Clarke, Samuel (1705). *A Discourse of Natural Religion.*

Cohen, G. A. (1996). 'Reason, humanity and the moral law'. In Korsgaard 1996b.

Cullity, Garret and Gaut, Berys (eds.) (1997). *Ethics and Practical Reason.* Oxford: Oxford University Press.

Cupitt, Don. (1980). *Taking Leave of God.* London: SCM Press.

—— (1995). *Solar Ethics.* London: SCM Press.

—— (1997). *After God: The Future of Religion.* London: Weidenfeld & Nicolson.

Dewey, John (1922). *Human Nature and Conduct.* New York: Henry Holt.

Donagan, Alan (1977). *The Theory of Morality.* Chicago: University of Chicago Press.

—— (1993). 'Moral Dilemmas, Genuine and Spurious: A Comparative Anatomy', *Ethics* vol. 104 (October 1993): pp. 7-21.

Dreier, James (1997). 'Humean Doubts about the Practical Justification of Morality' in Cullity and Gaut (1997): pp. 81-101.

Easwaran, Eknath (1986). *The Dhammapada.* London: Routledge & Kegan Paul.

Eliot, George (1860). *The Mill on the Floss,* ed. Gordon S. Haight. Oxford: Clarendon Press, 1980.

Epictetus, *Encheiridion.* Translated as *The Handbook of Epictetus* by Nicholas White. Indianapolis: Hackett (1983).

Flanagan, Owen (1990). 'Identity and Strong and Weak Evaluation', in *Identity, Character and Morality: essays in Moral Psychology* eds. Owen Flanagan and Amélie Oksenberg Rorty. Cambridge Mass.: MIT Press (1990).

Foucault, Michel (1997). *Ethics: Subjectivity and Truth.* Volume One of *The Essential Works of Michel Foucault,* ed. Paul Rabinow. London: Allen Lane The Penguin Press.

Frankfurt, Harry (1971). 'Freedom of the Will and the Concept of a Person', *Journal of Philosophy* vol. LXVIII, no.1. Reprinted in Frankfurt (1988).

—— (1988). *The Importance of what we care about.* Cambridge: Cambridge University Press.

Gibbard, Alan (1999). 'Morality as Consistency in Living: Korsgaard's Kantian Lectures'. *Ethics* vol. 110 (October 1999): pp. 140-164.

Glover, Jonathan (1977). *Causing Death and Saving Lives.* Harmondsworth: Penguin.

—— (1999) *Humanity: A Moral History of the Twentieth Century.* London: Jonathan Cape.

Habermas, Jürgen (1990a). 'Justice and Solidarity: On the Discussion Concerning Stage 6', in Wren ed. *The Moral Domain: Essays in the Ongoing Discussion between Philosophy and the Social Sciences.* Cambridge, Mass.: MIT Press.

—— (1990b). 'Morality and ethical life: does Hegel's critique of Kant apply to discourse ethics', in his *Moral Consciousness and Communicative Action.* Trans. Christian Lenhardt and Shierry Weber Nicholson. Cambridge, Mass.: MIT Press (1990).

—— (1996). 'On The Cognitive Content of Morality'. *Proceedings of the Aristotelian Society* (1996).

—— (1999). 'From Kant to Hegel and Back Again: The Move Towards Detranscendentalization'. *European Journal of Philosophy* vol. 7.2, (1999): pp. 129-157.

Hampton, Jean (1998). *The Authority of Reason.* Cambridge: Cambridge University Press.

Harris, George W. (1997). *Dignity and Vulnerability: Strength and Quality of Character.* California: University of California Press.

Hegel, G. W. F, (1977). *Phenomenology of Spirit,* trans. A. V. Miller. Oxford: Oxford University Press.

—— (1983). 'Jena Lectures on the Philosophy of Spirit', in *Hegel and the Human Spirit: A Translation of the Jena Lectures on the Philosophy of Spirit (1805-6) with Commentary,* ed. and trans. Leo Rauch. Detroit: Wayne State University Press.

Homer, *The Iliad,* trans. Martin Hammond (1987). London: Penguin.

Honneth, Axel (1995). *The Struggle for Recognition,* trans Joel Anderson. Cambridge: Polity Press.

Hundert, Edward M. (1989). *Philosophy, Psychiatry and Neuroscience: Three Approaches to the Mind.* Oxford: Clarendon Press.

James, William (1897). 'The Will To Believe'. Reprinted in James (2000).

—— (1902). *The Varieties of Religious Experience.* USA: Longmans, Green and Co.

—— (2000). *Pragmatism and Other Writings.* Harmondsworth: Penguin.

Joas, Hans (2000). *The Genesis of Values,* trans. Gregory Moore. Cambridge: Polity Press.

Jünger, Ernst (1929). *The Storm of Steel*: *From the diary of a German Storm-Troop Officer on the Western Front.* Translated by Basil Creighton. London: Chatto & Windus. Original German edition: *In Stahlgewittern,* 1921.

Kant, I. (1781). *Critique of Pure Reason* [*C1*]*,* trans. Norman Kemp Smith (1929). London: The Macmillan Press.

—— (1785). *Grounding for the Metaphysics of Morals* [*G*]*,* trans. James W. Ellington (2nd ed. 1994). Indianapolis: Hackett Publishing Company.

—— (1788). *Critique of Practical Reason* [*C2*]*,* trans. Lewis White Beck (3rd ed. 1993). New Jersey: Macmillan Publishing Company.

—— (1793). *Religion within the limits of reason alone* [*R*]*,* trans. Theodore M. Greene and Hoyt H. Hudson. New York: Harper & Brothers (1960).

—— (1793). 'On the Common Saying: "This May be True in Theory, but it does not Apply in Practice"' in *Kant's Political Writings* ed. Hans Reiss. Cambridge: Cambridge University Press. 2nd ed. 1991.

—— (1797). *The Metaphysics of Morals* [*MM*]*,* trans. Mary Gregor. Cambridge: Cambridge University Press (1991).

Keenan, Brian (1992). *An Evil Cradling.* London: Hutchinson.

Korsgaard, Christine (1983). 'Two Distinctions in Goodness', *The Philosophical Review,* vol. 92, no. 2 (April 1983): pp. 169-195. Reprinted in Korsgaard 1996a.

—— (1986). 'Kant's Formula of Humanity', *Kant-Studien* vol. 77, Jahrgang, Heft 2: pp. 183-202. Reprinted in Korsgaard 1996a.

—— (1985). 'Kant's Formula of Universal Law', *Pacific Philosophical Quarterly,* vol. 66, nos. 1&2 (January/April 1985): pp. 24-47. Reprinted in Korsgaard (1996a).

—— (1989a). 'Personal Identity and the Unity of Agency: A Kantian Response to Parfit', *Philosophy and Public Affairs* vol. 18, no. 2 (Spring 1989): pp. 101-132. Reprinted in Korsgaard 1996a.

—— (1989b). 'Morality as Freedom', in *Kant's Practical Philosophy Reconsidered* ed. Yirmiyahu Yovel. Dordrecht: Kluwer Academic Publishers: pp. 23-48. Reprinted in Korsgaard 1996a.

—— (1996a). *Creating the Kingdom of Ends.* Cambridge: Cambridge University Press.

—— (1996b). *The Sources of Normativity.* Cambridge: Cambridge University Press.

—— (1997). 'The Normativity of Instrumental Reason' in Cullity and Gaut eds (1997).

Laing, R D (1960). *The Divided Self.* London: Tavistock.

Leiter, Brian (1994). 'Perspectivism in Nietzsche's *Genealogy of Morals*' in Schacht ed. (1994).

Long, A. A. & Sedley D. N., (1987). *The Hellenistic Philosophers: volume 1.* Cambridge: Cambridge University Press.

MacIntyre, Alasdair ed. (1972). *Hegel: A collection of Critical Essays.* Garden City: Doubleday.

—— (1981). *After Virtue: A Study in Moral Theory.* London: Duckworth.

Margalit, Avishai (1996). *The Decent Society,* trans Naomi Goldblum. London: Harvard University Press

Murdoch, Iris (1992). *Metaphysics as a Guide to Morals.* Harmondsworth: Allen Lane The Penguin Press.

Nagel, Thomas (1996). 'Universality and the reflective self' in Korsgaard (1996b).

Nehamas, Alexander (1985). *Nietzsche: Life as Literature.* London: Harvard University Press.

Nietzsche, Friedrich (1882). *The Gay Science,* trans. Walter Kaufmann (1974). New York: Vintage Press.

—— (1886). *Beyond Good and Evil: Prelude to a Philosophy of the Future,* trans. R. J. Hollingdale (1973). Harmondsworth: Penguin.

—— (1887). *On the Genealogy of Morality: A Polemic,* trans. Carol Diethe (1994). Cambridge: Cambridge University Press.

—— (1889). *Twilight of the Idols,* trans. R. J. Hollingdale (1968). Harmondsworth: Penguin.

—— (1895). *The Anti-Christ,* trans. R. J. Holligdale (1968). Harmondsworth: Penguin.

Nussbaum, Martha (1986). *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy.* Cambridge: Cambridge University Press.

—— (1994), 'Pity and Mercy: Nietzsche's Stoicism' in Schacht ed. (1994).

—— (1997). 'Kant and Cosmopolitanism' in Bohman and Lutz-Bachmann eds. (1997).

Nussbaum, Martha and Sen, Amartya eds. (1993). *The Quality of Life.* New York: Oxford University Press.

O'Neill, Onora (1989). *Constructions of Reason.* Cambridge: Cambridge University Press.

Oppenheimer, Paul (1996). *Evil and the Demonic: A New Theory of Monstrous Behaviour.* New York: New York University Press.

Parker, Tony (1983). *The People of Providence: A Housing estate and some of its inhabitants.* London: Hutchinson.

—— (1990). *Life After Life: Interviews with twelve murderers.* London: Martin Secker & Warburg Ltd.

Perelman, C. and Olbrechts-Tyteca L. (1970). *La nouvelle rhétorique. Traité de l'argumentation.* Paris 1958, 2nd ed. Brussels 1970.

Prichard, H. A. (1912). 'Does Moral Philosophy Rest on a Mistake?', *Mind* vol. 21.

Putnam, Hilary (1982). 'Why Reason Can't Be Naturalized', *Synthese* vol. 52: pp. 1ff. Reprinted in Baynes, Bohman and McCarthy (1987) pp. 222ff.

Railton, Peter (1997). 'On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action', in Cullity and Gaut (1997): pp. 53-80.

Rehg, William (1994). *Insight and Solidarity: The Discourse Ethics of Jürgen Habermas.* London: University of California Press.

Ricoeur, Paul (1990). 'The Golden Rule: Exegetical and Theological Perplexities', *New Testament Studies* vol. 36, pp.392-7.

—— (1991). 'Love and Justice'. In *Radical Pluralism and Truth: David Tracy and the Hermeneutics of Religion,* ed. Werner G. Jeanrond and Jennifer L. Rike. New York: Crossroad, pp. 187-202. Reprinted in Ricoeur 1995.

—— (1992). *Oneself as Another* trans. Kathleen Blamey. Chicago: Chicago University Press.

—— (1995). *Figuring the Sacred* trans. David Pellauer, ed. Mark I. Wallace. Minneapolis. Fortress Press.

Rorty, Richard (1993). 'Human Rights, Rationality and Sentimentality' in *On Human Rights: The 1993 Oxford Amnesty Lectures,* ed. Susan Hurley and Stephen Shute. New York: Basic Books. Reprinted in Rorty 1998.

—— (1998). *Truth and Progress: Philosophical Papers volume 3.* Cambridge: Cambridge University Press.

Rudd, Niall (1998). *Cicero: The Republic and The Laws.* Oxford: Oxford University Press.

Scanlon, T. M. (1998). *What We Owe to Each Other.* USA: Harvard University Press.

Schacht, Richard ed. (1994). *Nietzsche, Genealogy, Morality: Essays on Nietzsche's Genealogy of Morals.* London: University of California Press.

Schopenhauer, Arthur [1841] (1995). *On the Basis of Morality.* Oxford: Berghahn Books.

Sidgwick, Henry [1907] (1962). *The Methods of Ethics.* 7th ed. Chicago: University of Chicago Press.

Sekida, Katsuki (1976). *Zen Training: Methods and Philosophy,* ed. A. V. Grimstone. New York: Weatherhill Publishers.

Silber, John R., (1960) 'The Ethical Significance of Kant's *Religion*' in Greene & Hudson's edition of *Religion within the limits of reason alone.*

Singer, Marcus George (1961). *Generalization in Ethics.* New York: Alfred Knopf.

Smith, Nicholas (1997). *Strong Hermeneutics: Contingency and moral identity.* London: Routledge.

Taylor, Charles (1964). *The Explanation of Behaviour.* London: Routledge and Kegan Paul.

—— (1972). 'The opening arguments of the phenomenology' in MacIntyre (1972): pp. 151-188.

—— (1977). "What Is Human Agency?" in *The Self: Psychological and Philosophical Issues*, Theodore Mischel, ed. Oxford: Blackwell pp. 103-35. Reprinted in Taylor (1985a) pp. 15-44.

—— (1985a). *Human Agency and Language: Philosophical Papers 1* Cambridge: Cambridge University Press.

—— (1985b). "Self-interpreting Animals," in Taylor 1985a: pp. 45-76.

—— (1985c). "The concept of a person," in Taylor 1985a.

—— (1987). "Overcoming Epistemology," in Baynes, Bohman and McCarthy (1987) pp. 464-88.

—— (1989). *Sources of the Self: The Making of the Modern Identity* Cambridge: Harvard University Press.

—— (1993). 'Explanation and Practical Reason' in Nussbaum and Sen (1993), pp. 208-31.

—— (1995). 'A very peculiar institution', in Altham and Harrison (1995)

Terkel, Studs (1972). *Working: People Talk About What They Do All Day and How They Feel About What They Do.* New York: The New Press.

—— (1992). *Race: How Blacks and Whites Think and Feel About the American Obsession.* New York: The New Press.

Velleman, J. David (1992). 'The Guise of the Good', *Nous* vol. 26: pp. 3-26.

—— (1996). 'The Possibility of Practical Reason', *Ethics* vol. 106: pp. 694-726.

—— (1997). 'Deciding How to Decide', in Cullity and Gaut (1997)

—— (1999). 'Love as Moral Emotion', *Ethics* vol. 109: pp. 338-374.

Wattles, Jeffrey (1996). *The Golden Rule.* New York: Oxford University Press.

Williams, Bernard (1985). *Ethics and the Limits of Philosophy.* London: Collins.

—— (1993). 'Nietzsche's Minimalist Moral Psychology', *European Journal of Philosophy,* vol. 1, no.1 pp.1-14. Reprinted in Schacht (1994).

Winnicott, Donald W (1949), 'The Mind and Its Relation to the Psyche-Soma'. In Winnicott 1958a.

—— (1958a). *Collected Papers: Through Paediatrics to Psycho-Analysis.* London: Tavistock Publications.

—— (1958b). 'The Capacity to be Alone'. *International Journal of Psycho-Analysis,* vol. 39, reprinted in Winnicott (1965).

—— (1960a). 'Ego Distortion in Terms of True and False Self' in Winnicott (1965), pp. 140-152.

—— (1960b). 'The Theory of the Parent-Infant Relationship'. *International Journal of Psycho-Analysis,* vol. 41 (1960). Reprinted in Winnicott (1965), pp. 37-55.

—— (1962). 'Ego Integration in Child Development', in Winnicott 1965, pp. 56-63.

—— (1965). *The Maturational Processes and the Facilitating Environment.* London: Hogarth Press Ltd.

—— (1969). 'The Use of an Object and Relating through Identifications'. *International Journal of Psycho-Analysis,* vol. 50 (1969). Reprinted in Winnicott (1971).

—— (1971). *Playing and Reality.* London: Tavistock Publications.