

ARISTOTLE AND AUGUSTINE

ON VOLUNTARY ACTION

AND FREEDOM AND WEAKNESS OF THE WILL

1. 'Positive' and 'Negative' Theories of Freedom
2. Compulsion, Duress, Persuasion, and Free Will
3. The Varieties of **a thesis submitted by**
4. Irrationality

Chapter Two: Freedom, AD 11-12, 20-21, 200-201, 202

1. Function **Timothy David John Chappell, M.A. (Oxon.)**
2. Aristotle's Epistemology

Chapter Three: Providence and Practical Reasoning

1. Providence
2. τὸ ἐκόντων, Action in accordance with the difference between them
3. Providence
4. Practical Reasoning and the 'Practical syllogism'
5. Conclusion **in partial fulfilment of the requirements**

Chapter Four: The Varieties for the Degree of

1. How to Solve the Problem of Aristotle
2. How not to Solve the Problem of Aristotle's

University of Edinburgh, January 1992

3. The Varieties of Aristotle



Table of Contents

<u>Acknowledgements</u>	7
<u>Declaration</u>	8
<u>Abstract of Thesis</u>	9
<u>Chapter One: Aristotle on the less than Fully Voluntary</u>	10
1. 'Positive' and 'Negative' Theories of Freedom	
2. Compulsion, Duress, Persuasion, and Free Action	
3. The Varieties of Ignorance	
4. Irrationality	
<u>Chapter Two: Freedom, Ability and Knowledge</u>	59
1. Function, Process and 'Ability to do Otherwise'	
2. Aristotle's Epistemology	
<u>Chapter Three: Proairesis and Practical Reasoning</u>	125
1. προαίρεσις	
2. τό ἐκούσιον, Action on προαίρεσις and the Difference between them	
3. βούλευσις	
4. Practical Reasoning and the 'Practical Syllogism'	
5. Conclusion	
<u>Chapter Four: The Varieties of Akrasia</u>	188
1. How to Solve the Problem of Akrasia	
2. How not to Solve the Problem of Akrasia: Aristotle's Account	
3. The Varieties of Akrasia	

Chapter Five: Freedom and the Voluntary

in Augustine's 'de Libero Arbitrio'.....243

1. Introduction

2. Book One

3. Book Two

4. Book Three

Chapter Six: Voluntariness and Responsibility

in Augustine.....301

1. Introduction

2. The Linking of Voluntariness and Responsibility

3. Two of Augustine's Conditions of Voluntary Action

4. 'Ignorantia' and 'Difficultas'

5. From the Earlier to the Later Theory

6. Conclusion

Chapter Seven: 'Voluntas' and the Voluntary.....329

1. Introduction

2. 'Voluntas': 'Voluntary Action' or 'Volition'?

3. The Nature of 'Voluntas': Two Requirements

4. Ability to Do Otherwise

5. The Cause(s) of 'Voluntas' Again

6. A Reflexive 'Voluntas'?

Chapter Eight: The Good Will and the Good Life.....355

1. Introduction

2. Practical reason and Practical Wisdom

in Augustine

3. Aristotle and Augustine on the Directedness of Action

4. 'Felicitas'

5. Good Will and the Order of the World

Chapter Nine: 'Mala Voluntas' and the Mystery
of Evil.....391

1. Introduction

2. 'Mala Voluntas'

3. An Incomplete Account?

4. A *Necessarily* Incomplete Account?

5. Manichaeism and 'Privatio Boni'

6. Incompleteness Again

7. Augustine and Voluntarism

8. From Augustine's Rationalism to
 Augustinian Voluntarism

9. Conclusion

Bibliography.....442

Tables:

Table One: The Hierarchy of Nature.....85

Table Two: The Sixty Four Permutations of the Conditions
 of Akrasia.....242

In Memoriam

Gillian Patriciae Chappell

30.4.1937-

13.12.1989

Quia fecisti nos ad Te

et inquietum est cor nostrum

donec requiescat in Te

(Peter Seewald, *Philosophy of the Natural Academy*, 1962)

'By liberty, then, we can only mean a power of acting or not acting, according to the determinations of the will; that is, if we choose to remain at rest, we may; if we choose to move, we also may. Now this hypothetical liberty is universally allowed to belong to everyone who is not a prisoner and in chains. Here, then, is no subject of dispute.'

(David Hume: *An Enquiry Concerning Human Understanding* (1775), §VIII.1)

'It does not seem to be self contradictory to suppose that [the acceptance of determinism could lead to the decay or repudiation of participant reactive attitudes]... But I am strongly inclined to think that it is, for us as we are, practically inconceivable. The human commitment to participation in ordinary inter-personal relationships is too thoroughgoing and too deeply rooted for us to take seriously the thought that a general theoretical conviction might so change our world that, in it, there were no longer any such things as interpersonal relationships as we normally understand them.'

(Peter Strawson, *Proceedings of the British Academy*, 1962)

Acknowledgements

This thesis has been completed under the supervision of Professor James Mackey (principal supervisor) and Mr. David Wright (second supervisor). I wish to thank both gentlemen for the benefit of their wisdom, scholarship and advice. I have also learnt much, in various more or less formal contexts, from a number of other interlocutors; in particular, from Drs. Lubor Velecky and David Pugmire of Southampton University, my first postgraduate supervisors; also from Dr. Dory Scaltsas, Mr. Stephen Priest, and Mr. William Charlton, of Edinburgh University; from Father Fergus Kerr, O.P., and Br. Hayden Ramsay, O.P., of Blackfriars, Edinburgh; from Dr. Christopher Martin, of Glasgow University; and from Dr. Roger Crisp, of St. Anne's College, Oxford. Nonetheless, of course, my work as here presented remains entirely my own responsibility, and none of these who have helped me is any degree answerable for its deficiencies.

In another sense, I am indebted too to the Whitefield Institute, Oxford, for their generous support, both financial and moral, over the last three years. I am also grateful to my parents for their love, support and encouragement during my doctoral studies.

But my biggest debt of all is to my wife Claudia, without whom this thesis would never have been begun, let alone finished.

Timothy Chappell

Wolfson College, Oxford

January 1992

Declaration

All the material contained in this thesis is my own work, and results from original research carried out by me in Southampton, Edinburgh and Oxford between October 1988 and January 1992.

The only exception to this is constituted by those passages of the thesis in which I have quoted from Aristotle, Augustine, the secondary literature, and elsewhere. The arrangement of these quotations is, however, of my own devising.

At the time of writing, none of the work by me which is contained herein has been published in any other form than as here presented.

Signed:

January 1992

Thesis Abstract

Aristotle's remarks on free will suggest, not so much an argument for the existence of free will, as an account of its nature. This account depends on his making no hard distinction between what we call 'free action' and 'voluntary action'. For him, these would be interchangeable terms. The Aristotelian can, then, point out that, if we give up our belief in free will, we must give up many other natural beliefs too. In particular, we must stop believing in voluntary action.

There are, in Aristotelian terms, three conditions (not two, as Aristotle himself evidently supposed), which any behaviour must satisfy to count as free/voluntary action. The behaviour (i) must not be compelled, but must be performed by the agent's own power and desire; (ii) must not be done in ignorance, but must be action on relevant knowledge; and (iii) must not be irrational, but must result from the combination of the agent's own power and desire with the agent's relevant knowledge. (i) leads me to discuss Aristotle's account of what he calls *kineseis*; (ii) leads me into epistemology; (iii) into an account of Aristotle's theory of *proairesis* and practical reasoning as the cause of voluntary action.

One problem for Aristotle's account of the causation of voluntary action is posed by *akrasia*, deliberate choice of what I sincerely believe I should not choose. This seems to be voluntary action which is not caused as Aristotle says voluntary action should be. But the three conditions of voluntary action which I say Aristotle should be committed to can be used to show that the existing forms of *akrasia* make no counter example to Aristotle's theory, but rather an interesting adjunct to it.

My study of Augustine's theory of freedom begins with a survey of a crucial text, the *de Libero Arbitrio* (Ch.5). I then apply an analogous schema to that found in Aristotle. Augustine too depends on the idea that to analyse free action is to analyse voluntary action; he also equates these two types with responsible action. He too believes (i) that ignorance usually makes for involuntariness, and (ii) that there can be no voluntary action which is compelled or which the agent could not have done otherwise. In his later works, these doctrines are often obscured by his interest in original sin and predestination (neither of which topics, be it noted, are focuses of this thesis). But they remain his doctrines.

Does Augustine have (iii) any doctrine that voluntary action must be rational? While he does not develop any theory of practical reasoning like Aristotle's, he does develop a theory of practical wisdom. It is an essential feature of all human desire, and hence of all voluntary action, that it aims at happiness, which properly understood is identical with possession of The Good, i.e. of God. From this Augustine draws the conclusion that, to explain any behaviour as a voluntary action or choice, it is necessary and sufficient to specify some good at which it is to be understood as aiming.

This sets up for Augustine a problem analogous to Aristotle's problem about *akrasia*. How is a voluntary choice of evil explicable? Augustine's reply is that human desires have been disordered by the Fall, and so we often choose, not evils *per se*, but lesser goods than we ought. But this prompts the question: How is a first voluntary choice of evil explicable? Augustine's reply is simply that it is not. Since a voluntary action or choice must be explained by reference to some good at which it aims, a voluntary choice of evil *per se* cannot be explained at all. This does not mean that there was no voluntary choice of evil; but it does mean that, in principle, that choice is inexplicable- a mystery. Thus Augustine, unlike Aristotle, in this one exceptional case (but in no others) affirms that there can be genuinely voluntary action which is not, in the relevant sense, rational.

Aristotle on the Less than Fully Voluntary

1. 'Positive' and 'Negative' Theories of Freedom
2. Compulsion, Duress, Persuasion, and Free Action
3. The Varieties of Ignorance
4. Irrationality

'While it has been the tradition to present [freedom] as the positive term requiring elucidation, there is little doubt that to say we acted "freely"... is to say only that we acted not *un*-freely... Like "real", "free" is only used to rule out the suggestion of some or all of its recognised antitheses... Aristotle has often been chidden for overlooking "the real problem": in my own case, it was when I began to see the injustice of this charge that I first became interested in excuses.'

Second, Austin argues that free action is not a 'positive term' (Austin, 1956-7) but (in effect) a paradigm case. When we talk about action, says Austin, we mean free action. Austin explicitly enumerates 'all the ways in which each

1. 'Positive' and 'Negative' Theories of Freedom

Austin's famous remarks suggest a view of Aristotle's theory of freedom to act which is worth taking seriously on at least two counts. First, Austin takes it that 'freedom' is not something

arcane and mysterious, but as near and familiar to us as voluntary action. Thus, developing an adequate theory of the nature of *voluntariness* is both necessary and sufficient for developing an adequate theory of the nature of *freedom to act*.

But a *caveat* is necessary to the stating of this methodological assumption. This *caveat* is that an adequate *theory of the nature of freedom to act* is not a *proof of the existence of freedom to act*. Austin seems to obscure this point when he says that 'In examining all the ways in which it will not do to say simply "X did A", we may hope to dispose of the problem of freedom' (*loc.cit.*). This thesis does not aim to show that there *is* free will, so much as to show, like Strawson in 'Freedom and Resentment', how many other beliefs would have to be abandoned along with a belief in free will, *if* we were to abandon that belief. In particular, as already suggested, we would have to stop believing in the existence of any genuinely voluntary action.

Second, Austin argues that free action is not a 'positive term requiring elucidation', but (in effect) a *paradigm case*. When we talk about action *sans phrase* (he says), we *mean* free action. Austin explicitly equates 'all the ways in which each action may not be free' with 'all the ways in which it will not do to say simply "X did A"' (Austin, *loc. cit.*). For him, qualifications are only added when the action is in some way or other *less than* fully free. Therefore we arrive at our definition of free action, according to Austin, by a kind of subtraction. If

you imagine a case of human action, then take away from it all the possible factors that could militate against its being an ordinary case of human action, what you are left with is *typical* action. And (he argues) typical action, as normally understood, *just is free action*.

I shall call any theory of this Austinian sort a *negative* theory of free action. Austin rightly points out that the claim that Aristotle himself has a negative theory of free action is readily supportable from the ethical writings:

ὄσα μὲν οὖν ἐφ' ἑαυτῷ ὄν μὴ πράττειν πράττει μὴ ἀγνοῶν καὶ δι' αὐτόν, ἐκούσια ταῦτ' ἀνάγκη εἶναι, καὶ τὸ ἐκούσιον τοῦτ' ἐστίν. (EE 1225b7-10)

ὄντος δ' ἀκουσίου τοῦ βίᾳ καὶ δι' ἀγνοίαν, τὸ ἐκούσιον ὁξείειεν ἂν εἶναι οὐ ἢ ἀρχὴ ἐν αὐτῷ εἰδῶτι τὰ καθ' ἕκαστα ἐν οἷς ἢ πρᾶξις. (NE 1111a22-24)

My first job in this thesis is to give an account of this negative theory, which is to say an account of the cases which for Aristotle constitute what I call 'the less than fully voluntary'.

But first, a comment on the very idea of a purely negative theory of free action such as Austin proposes. I agree with Austin that, if we want to understand what 'freedom to act' means, the cases where it is impeded are revealing. But how

simple a matter *is* it to 'say simply "X did A"'? Does that fact of x's doing A, once arrived at, really require no further explanation? Iris Murdoch noted that 'Oxford philosophy' (by which I think she meant chiefly Austin himself) 'has developed no serious theory of motivation' (Murdoch 1970, p.53). Is there really *nothing* of interest or importance to be said which might contribute towards the development of a *positive* theory of free action? If it turns out that there is anything non-trivial and plausible to say towards such a positive account, then it follows that a negative theory alone is incomplete.

And indeed (as I will argue), besides his undoubted interest in what free action *isn't*, Aristotle also has plenty to say about what free action *is*. Aristotle does have a well developed negative theory about what kinds of opposing factors count as impairments or suspensions of normal freedom to act. But he has an even better developed positive theory about what typical action consists in, what it is for an agent to have a normal freedom to act. It is unfortunate that Aristotle's positive doctrine is frequently overlooked (as here, I suspect, by Austin), or else treated as if it had no bearing on the issue of freedom.

In this thesis I present both negative and positive aspects of Aristotle's theory of free action. It will turn out that the negative and positive conditions of free action which we can supply from Aristotle are corollaries of each other.

But- against Austin- it is Aristotle's positive theory which is prior and more fundamental. For to define and enumerate the kinds of obstacles that there are to free action is to list the varieties of ill formed free action that there are. But this can only prompt the prior question: 'What is the *normal* "form" of free action?'. And Aristotle does believe- unlike, perhaps, Sartre (v. Sartre 1949)- that we can answer this, can define freedom of action by giving it a certain structure. His positive theory describes that structure.

But I start with the negative theory. What, for Aristotle, must *not* be the case regarding an agent, if he is to do a free or voluntary action? The passage normally considered vital has been cited already:

ὄντος δ' ἀκουσίου τοῦ βίῃ καὶ δι' ἄγνοιαν... (NE 1111a22)

On the basis of these undeniably very explicit words, it is standardly assumed that Aristotle believes there to be two, and only two, negative conditions of free or voluntary action:

- (i) Voluntary action must not be compelled.
- (ii) Voluntary action must not be done in ignorance.

But does NE 1111a22 give us all the kinds of involuntariness that there are? In §4 I shall argue that Aristotle himself sometimes hints at a separate third negative condition, which, in the logic of his own position, really requires further

development than he gives it. This third negative condition, the sense of which will be explained in §4, is this:

(iii) Voluntary action must not be irrational behaviour.

Two comments are in order before I begin exegesis of the evidence for these conditions:

A. I use 'behaviour' to include all human performances, movements and physical manifestations, without prejudice to the question whether such performances count either as action or as free action.

B. As Austin noted, most of Aristotle's account of the negative conditions of the voluntary is concerned with *exculpation* from responsibility for *bad* deeds, with blame rather than with praise. As I have noted, this does not necessarily mean that Aristotle is *only*, or even *mainly*, concerned with excuses. Whatever he says about exculpation from blame I believe applies, *mutatis mutandis*, to exclusion from praise.

2. Compulsion, Duress, Persuasion, and Free Action

It is an obvious intuition that I cannot act freely or voluntarily if I am under compulsion:

τὸ τε γὰρ βίαιον ἀκούσιον καὶ τὸ ἀκούσιον πᾶν βίαιον φαμέν.

(EE 1224b11-12)

δοκεῖ δὴ ἀκούσια εἶναι τὰ βία... γινόμενα... (NE 1109b35)

The distinctions which Aristotle goes on to draw concerning compulsion and its relation to the voluntary are all qualifications or dilutions of the meaning of βία (or ἀνάγκη). What, then, is the basic theme on which the variations are to be played: what is to count as a case of straightforward compulsion?

2a. Straightforward Compulsion: Within and Outside the Agent

Aristotle opens his discussions of voluntariness and compulsion in the EE with a plausible generalisation: What something is *compelled* to do is whatever is *unnatural* (παρὰ φύσιν], EE 1224a20) for it to do. Correspondingly, what something does freely or spontaneously is whatever is *natural* (κατὰ τὴν φύσει καὶ τὴν καθ' αὐτὰ ὁρμὴν, EE 1224a18) for it to do. This opposition παρὰ φύσιν/ κατὰ φύσιν can be given fairly precise sense. The movement or process (κίνησις) which is *natural* to x, is whatever κίνησις originates within x. The κίνησις which is *unnatural* to x, is whatever κίνησις originates outside x:

βίαιον δὲ οὐ ἢ ἀρχὴ ἐξωθεν, τοιαύτη οὔσα ἐν ᾧ μηδὲν συμβάλλεται ὁ πράττων ἢ ὁ πάσχων... (NE 1110a1-3)

ὅταν μὲν γὰρ τι τῶν ἔξωθεν παρὰ τὴν ἐν αὐτῷ ὀρμὴν κινῆ ἢ ἡρεμίζῃ, βίᾳ φαμέν, ὅταν δὲ μή, οὐ βίᾳ. (EE 1224b7-9)

Three examples of this kind of compulsion. (i) Aristotle reports an interesting linguistic fact: that the movement of a stone thrown upwards, or of a flame made to burn 'upside down', were called 'forced' movements by Greeks of his time:

καὶ γὰρ τὸν λίθον ἄνω καὶ τὸ πῦρ κάτω βίᾳ καὶ ἀναγκαζόμενα φέρεσθαι φαμέν. (EE 1224a17)

This kind of movement is contrasted by Aristotle with the opposite movements (downwards for a stone, upwards for a flame), the movements κατὰ τὴν φύσει καὶ τὴν καθ' αὐτὰ ὀρμὴν (EE 1224a18). These (he reports) were not called forced movements (a19)- though admittedly they were not called *voluntary* movements either (οὐ μὴν οὐδ' ἐκούσια λέγεται, ἀλλ' ἀνόνημος ἢ ἀντίθεσις, a20). These movements are 'natural' for these things because they are the movements which, according to Aristotle's physics, these things will display if left to their own devices, the movements which originate within these things ὅταν δ' ἔσωθεν ἢ ἀρχή, οὐ βίᾳ (EE 1224b15).

This general point is applicable to human beings, whose 'natural' movement, as we shall see further in Ch.2, is voluntary action. Thus Aristotle tells us that an 'origin (ἀρχή) outside the agent' means an ἀρχή which is τὴν παρὰ τὴν ὀρμὴν ἢ ἐμποδίζουσαν ἢ κινουσαν ἀνάγκην (EE 1224b13). For example, (ii),

...ὥσπερ εἴ τις λαβὼν τὴν χεῖρα τύπτοι τινὰ ἀντιτείνοντος καὶ τῷ βούλεσθαι καὶ τῷ ἐπιθυμεῖν. (EE 1224b13-15; cp. NE 1135a27)

Where I cannot prevent my limbs being utilised for another's purposes, I am straightforwardly compelled to perform whatever movements that other wishes. But the impetus to action (ὄρμη) does not arise within me. Hence I am not responsible for whatever is so done, because in such a case I do not act voluntarily.

(iii) In the NE there is the more famous example of the captain in his ship: ...οἶον εἰ πνεῦμα κομίσει ποὶ ἢ ἄνθρωποι κύριοι ὄντες (NE 1110a3). The captain is not responsible for the movements of his ship unless they are the results of his own ὄρμαι: that is, unless they are under his control, which they are not if typhoons or brigands commandeer his sails. He too is straightforwardly compelled, and therefore exonerated.

2b. Non-Compulsions from Outside

The plausible generalisation is, then, that I am responsible for all my behaviour which originates within me, but for none of it which originates outside me. However, as before, this generalisation is only a starting point for Aristotle. He thinks that, before it can be accepted (if it can), this statement should be tested against obvious objections, some more compelling than others. This procedure will allow us both to elaborate the original rule of thumb, and to note any exceptions to it.

The first objection to the rule of thumb is this. If compulsions arise only without and free actions only within, why not say that I am compelled (e.g.) by *what is pleasant* (perhaps in the case of an unrestrained or wicked action), or by *what is noble* (perhaps in the case of a virtuous action)?

Aristotle's response is as follows. Of course it is true that 'what is pleasant' and 'what is noble' are things- states of affairs conceived by me as desirable goals- outside me. But one cannot legitimately move from the rule of thumb to the claim that *any* outside influence on my behaviour, since it is outside me, must be a compulsion. To do so would destroy the very idea of a contrast between the voluntary and the involuntary which Aristotle is trying to establish:

εἰ δὲ τις τὰ ἡδέα καὶ τὰ καλὰ φαίη βίαια εἶναι (ἀναγκάζειν γὰρ ἔξω ὄντα), πάντα ἄν εἴη οὕτω βίαια. τούτων γὰρ χάριν πάντες πάντα πράττουσιν. (NE 1110b9-11)

It may be true that when, say, I seduce my neighbour's wife, I act on an external influence, namely her beauty. But does this *look* like a case of compulsion (which is supposed to be unpleasant, b12)? Since I am clearly not physically pained by the experience, but the opposite (b12-13), doesn't the plea of compulsion just look like a rather flimsy pretext (b14-15)?

More importantly, we should distinguish (although Aristotle does not, explicitly) between something's being an *influence* on

an agent's behaviour, and its being the ἀρχὴ of that behaviour. The seduction case is not assimilable to the case in which the sea captain finds himself, because the ἀρχὴ of the sea captain's movement is the winds (or the pirates). If the captain can do anything at all, he may hope, at best, to *influence* his ship's movements, but (as the case is described) he cannot be the ἀρχὴ of those movements. In the seduction case, as I have described it, it is the other way round. The ἀρχὴ of my behaviour is not the lady's beauty, but my desire to seduce her. Her beauty is an *influence* on, and no doubt a necessary condition for, my behaviour; but it is not the ἀρχὴ of that behaviour. Hence my behaviour's ἀρχὴ is *not*, in truth, outside me, and so cases like this provide no counter-example to Aristotle's rule of thumb.

A note is needed here on hedonism and the voluntary in Aristotle. Some recent writers, such as Kenny (1979), have taken Aristotle to argue quite seriously, at least sometimes, for an identification of the pleasant and the voluntary. But I doubt that the hedonistic strand in Aristotle's theory of the voluntary is more than a *Protagoras*-inspired distraction from the main events.

Aristotle does, sometimes, seem inclined to hedonism of this sort (e.g. NE 1110b12-13: καὶ οἱ μὲν βίᾳ καὶ ἄκοντες λυπηρῶς, οἱ δὲ διὰ τὸ ἡδὺ καὶ καλὸν μεθ' ἡδονῆς). But three possible theses should be distinguished here. From the strongest to the weakest: (i) When I make a voluntary choice, my criterion is some 'hedonistic calculus'. (ii) My actions and experiences (πράξεις

and νόθη) can be assessed as more or less voluntary in proportion to how much pleasure or pain they give me. (iii) A discomfort (whether great or small) is always occasioned in me by something's happening to me against my will; a pleasure (whether great or small) is always occasioned by something's happening to me as I choose.

All three theses seem to me both false and counter to the general tendency of Aristotle's philosophy. (i) is clearly not compatible with Aristotle's belief that προᾶξις is motivated not by desire, nor by reasoning about the best course of action, but by desire *plus* reasoning about the best course (where 'best' does not just equal 'most pleasant') (NE 1139a18-b6). It follows that an agent who πράττει will not necessarily always do what gives him the most pleasure, at least not in any non-trivial sense. (ii) also is unpromising, (a) because I obviously can choose, quite voluntarily, many things that I know will give me no pleasure at all, and indeed much pain, such as to die horribly for my country; and (b) because 'X is a pleasant experience which is happening to me against my will' is a perfectly intelligible remark, as anyone who has been tickled can testify. If (iii) means that there is always the *same* experience of pleasure (pain) when what happens to me is voluntary (involuntary), then Wittgenstein's arguments against the possibility of reidentifying any private experience apply (Wittgenstein 1967). Anyway, whether (iii) means this or just the weaker claim that there is always *some* experience of pleasure (pain) when what happens to me is voluntary (involuntary), the claim seems empirically dubious, and

not very interesting even if true. For— for obvious reasons— no such experience could provide any *criterion* of voluntariness. Nor, in fact, does Aristotle ever seriously threaten to use the pleasantness of voluntary actions as such a criterion. The truth is that the identification of the pleasant and the voluntary, for Aristotle, is no more than a δόξα (δόξει δ' ἂν πᾶν τὸ κατ' ἐπιθυμίαν ἐκούσιον εἶναι, EE 1223a28), and a δόξα which he rejects (ὅτι μὲν τοίνυν οὐκ ἔστι τὸ ἐκούσιον τὸ κατὰ ὄρεξιν πράττειν οὐδ' ἀκούσιον τὸ παρὰ τὴν ὄρεξιν φανερόν, EE 1223b37–38).

(Incidentally, it follows at once from this that Aristotle ought to say that the sense in which actions under duress are involuntary is not as important as the sense in which they are voluntary. See §2d below.)

2c. Compulsions from Within

So the rule of thumb survives the first objection, if it is modified to say that 'Voluntary action is behaviour that finds its ἀρχή, and not just some of the influences on its occurrence, within the agent; compelled behaviour is behaviour that finds its ἀρχή, and not just some of the influences on its occurrence, outside the agent'. But what if the objector now goes on the other tack? He has attempted to argue that no behaviour truly originates within the agent, and hence that no behaviour is voluntary. He may now attempt to argue a second objection: that

even behaviour which does truly originate within the agent is not necessarily voluntary.

This is a problem which Aristotle notes at EE 1224a23 ff., drawing a contrast between things like stones and things like human beings. Things like stones *either* undergo 'natural' movement (according to the $\delta\rho\mu\eta$ which is $\acute{\alpha}\rho\chi\eta$ within them) *or* 'unnatural' movement (according to some other $\delta\rho\mu\eta$, outside them). For things like human beings, however, this simple contrast is not available: $\acute{\epsilon}\nu\ \mu\acute{\epsilon}\nu\ \tau\omicron\iota\varsigma\ \acute{\alpha}\psi\acute{\upsilon}\chi\omicron\iota\varsigma\ \acute{\alpha}\pi\lambda\eta\ \eta\ \acute{\alpha}\rho\chi\eta$, $\acute{\epsilon}\nu\ \delta\acute{\epsilon}\ \tau\omicron\iota\varsigma\ \acute{\epsilon}\mu\psi\acute{\upsilon}\chi\omicron\iota\varsigma\ \pi\lambda\epsilon\omicron\nu\acute{\alpha}\zeta\epsilon\iota$ (EE 1224a28). The simple contrast between $\acute{\alpha}\rho\chi\eta$ outside and $\acute{\alpha}\rho\chi\eta$ within is not available for $\tau\acute{\alpha}\ \acute{\epsilon}\mu\psi\acute{\upsilon}\chi\alpha$, because at least some 'animate creatures' (if that is how to translate the phrase) do not have one and only one kind of $\acute{\alpha}\rho\chi\eta$ within them: they have two, $\lambda\acute{\omicron}\gamma\omicron\varsigma$ and $\delta\rho\epsilon\chi\iota\varsigma$ (1224a26). This is the case, at any rate, with humans ($\acute{\epsilon}\nu\ \delta'\ \acute{\alpha}\nu\theta\rho\acute{\omega}\pi\omega\ \acute{\epsilon}\nu\epsilon\sigma\tau\iota\nu\ \acute{\alpha}\mu\phi\omega$, a28). (Aristotle seems unsure whether or not to say the same of other animals, drawing first a distinction between the 'animate' as having more than one $\acute{\alpha}\rho\chi\eta$ in them and the 'inanimate' as having only one, and then a second distinction between humans as having more than one $\acute{\alpha}\rho\chi\eta$ in them and the rest, both animals and objects, as having only one.)

Can an agent's behaviour genuinely have its $\acute{\alpha}\rho\chi\eta$ within that agent, and yet be not voluntary action but compelled behaviour? Aristotle's answer to this in the NE is markedly different from that in the EE. In the NE, his position is quite clearly that any $\delta\rho\mu\eta$ which originates in me is entirely my responsibility:

ἔοικε δὴ τὸ βίαιον εἶναι οὐ ἕξωθεν ἢ ἀρχή, μηδὲν συμβαλλομένου τοῦ βιασθέντος. (NE 1110b16-17)

In the NE, the only true compulsion is external compulsion, that which is applied to me against my wishes by other agents or outside forces. It is a sufficient criterion of voluntariness that the agent should be the ἀρχή of his action, that the action should originate in him. On this criterion of compulsion, the question of whether I can *help* obeying the pull of my own desires towards certain objects is strictly irrelevant. Thus οὐ καλῶς λέγεται ἀκούσια εἶναι τὰ διὰ θυμὸν ἢ ἐπιθυμίαν (1111a24-25).

However, if we accept the EE's point about there being more than one kind of ἀρχή in humans (as the Aristotle of the NE undoubtedly did: NE 1102a29 ff.), a more tolerant approach is suggested, which indeed, in the EE, Aristotle adopts; most explicitly at EE 1225a25-33. Here Aristotle virtually contradicts the teaching of NE 1111a24-25. He compares those who act under desire (ἐπιθυμία) with those who are possessed by spirits (ἐνθουσιῶντας):

οὐ φομεν ἐφ' αὐτοῖς εἶναι οὐτ' εἰπεῖν δ' εἶπον οὐτε πράξαι δ' ἔπραξαν. ἀλλὰ μὴν οὐδὲ δι' ἐπιθυμίαν... (EE 1225a30-31)

In the case of desire-led action, too, it now appears, it can sometimes be true that we can lack control, either over some of our thoughts and emotions (διάνοιαί τινες καὶ πάθη, a32), or over the actions resulting from our thoughts and reasonings (κατὰ τὰς

τοιαύτας διανοίας καὶ λογισμούς, a33). In what cases, then, can this loss of control of one's beliefs, wants or behaviour be said to occur? For unless we can give a clear answer to that question, this excuse of being overpowered by desire will be available to every scoundrel for every indulgence. Aristotle remarks that the crucial question (εἰς δ' ἀνάγεται ὅλον, a26) is that of what a nature is able to bear (δ' ἢ αὐτοῦ φύσις οἷα τε φέρειν, a27). But what a person's nature can and cannot bear is not up to him (a28). Hence (in Philolaus' words), εἶναί τινας λόγους κρείττους ἡμῶν (a34).

So it seems that, in the EE at least, Aristotle does allow for a breach of his rule of thumb: there *can*, on this account, be internal compulsions. (The only place in the NE where Aristotle definitely entertains the idea that internal compulsion can diminish voluntariness is NE 1148a18-22.) Action on such a compulsion will contrast both with action on an external compulsion (where the ὄρχη, as we have seen, is not within the agent); and also with action on an internal urge which, however, does not count as a compulsion (where there will be a conscious assent, on the agent's part, to acting in this way: v. EE 1225a26-27, δ' δὲ μὴ οἷα τε μηδ' ἐστὶ τῆς ἐκείνου φύσει ὀρέξεως ἢ λογισμοῦ, οὐκ ἐφ' αὐτῷ).

The criterion of an internal compulsion, then, is given by the answer to the question 'Is this urge too strong for *this nature* to resist?' (EE 1225a27). This, of course, is a question to which (even given the same strength of urge) the answer will

vary from one nature to another. Hence the assignment of responsibility and exculpation, in cases where a plea of internal compulsion is entered, will be a subtle matter. But this subtilising of the account need not mean that *carte blanche* has been given to every scoundrel for every indulgence. It might, rather, seem an advantage of the EE account over the NE account that this kind of being overpowered is here allowed as an excuse. (It is curious, incidentally, that Kenny (1979), which *inter alia* is far and away the most distinguished recent commentary on the EE, makes no mention of this important point of difference between the EE and the NE. Especially if its allowance of internal compulsion is considered to be an advantage of the EE's account over the NE's, one might have thought that Kenny would want to stress this point.)

The rule of thumb, then, needs to be modified again, to read: 'Voluntary action is behaviour that finds its ἀρχή, and not just some of the influences on its occurrence, within the agent *and to which the agent consents without being overpowered*; compelled behaviour is *either (i) behaviour that finds its ἀρχή, and not just some of the influences on its occurrence, outside the agent; or else (ii) behaviour that finds its ἀρχή within the agent, but to which the agent does not consent because he is overpowered*'.

2d. Duress (External)

This, I think, completes Aristotle's remarks on compulsion

strictly so called. But he does have something to say about an obvious close relation of compulsion, which I will call 'duress':

λέγονται δὲ κατ' ἄλλον τρόπον βίᾳ καὶ ἀναγκασθέντες πράξει...
ὅταν πράττωσιν ὁ καὶ λυπηρὸν καὶ φαῦλον ὑπολαμβάνουσιν, ἀλλὰ μὴ
τοῦτο πράττουσιν πληγαὶ ἢ δεσμοὶ ἢ θάνατοι ὄσιν. (EE 1225a3-6)

Compulsion is *physical* necessity; duress is *practical* necessity. Action under duress is exemplified by action under a threat, like the tyrant's threat to one's family if one does not comply with his wishes (NE 1110a6-7), or the threat perceived from the storm by the sailor who jettisons the cargo (NE 1110a8-9). Aristotle says that there is some doubt about whether action under duress is voluntary or involuntary (NE 1110a7-8), and indeed his own remarks seem to illustrate this ἀμφισβήτησις.

In the NE, Aristotle begins with puzzles and moves towards a simpler account. He begins by calling actions under duress μικτὰ (NE 1110a11), half voluntary and half involuntary. His reason for calling actions under duress involuntary is, apparently, to do with considerations about second order choice. It is clear that no one would *choose to choose* between having their relatives slaughtered and cooperating in a tyrant's atrocities: no one would voluntarily enter a situation where these were the only options on offer for voluntary choice. Neither option is the kind of thing that one would ever choose to do ἀπλῶς, considered in its own right. Hence action under duress is (he says), in one sense, 'perhaps (ἴσως) involuntary':

έκούσια δὴ τὰ τοιαῦτα, ἀπλῶς δὲ ἴσως ἀκούσια· οὐδεὶς γὰρ ἂν ἔλοιτο καθ' αὐτὸ τῶν τοιούτων οὐδέν. (NE 1110a18-20)

On the other hand, he goes on to say, actions under duress can also be seen as voluntary: (i) because they are chosen (αἴρεται, a13) *at the time they are done* (τότε ὅτε πράττονται), even if not at any other time. But it is the chosenness or otherwise of an action at the time it is done which counts for its (in)voluntariness (a14). And (ii) because the rule of thumb applies to them: the ἀρχὴ of an action under duress is within the agent. Hence such actions are μικταὶ, although εὐόκασι δὲ μᾶλλον ἔκουσίοις (NE 1110a12).

In the EE, by contrast, Aristotle begins with a simpler account and introduces more and more difficulties for it. (This contrast, incidentally, seems rather good evidence for dating at least this part of the EE *earlier* than the corresponding part of the NE. If an early essay on a subject starts well then runs into perplexities, it is natural when returning to the topic to start by re-examining those perplexities.) At first he seems to be in much less doubt than in the NE about the status of actions under duress. He does not start by calling them μικταὶ or raising considerations about second order choice. He simply says that all actions under duress must be voluntary, because in every case of duress it is true that ἔξεστι γὰρ μὴ ποιεῖν ἀλλ' ἐκεῖνο ὑπομεῖναι τὸ νόθος (EE 1225a7-8). However, unfortunately, he then goes on to complicate this picture, in three ways.

(i) If (he says) it was up to the agent (ἐφ' αὐτῷ, a10) to avoid or not avoid the situation of choice under duress, then the agent is responsible for being in that situation, and so his choice, however painful, is voluntary (EE 1225a9-12). If, however (a12-14), he had no chance of avoiding that situation, then one may say that his action is under compulsion of a sort (βία πῶς, a13), though not full-blown (ἀπλῶς) compulsion. (Compare NE 1110a18's ἴσως ἀκούσια.)

(ii) Furthermore, the agent's action under duress is now said to be less than fully voluntary because he does not choose the action which he does *for its own sake*, but for some further purpose, e.g. saving his family's lives: οὐκ αὐτὸ τοῦτο προαιρεῖται ὁ πράττει, ἀλλ' οὐ ἕνεκα (EE 1225a14). (This remark seems to bring out another side of what Aristotle said at NE 1110a19-20: οὐδεὶς γὰρ ἂν ἔλοιτο καθ' αὐτὸ τῶν τοιούτων οὐδέν.)

(iii) Finally, Aristotle simply contradicts EE 1225a7-8: οὕτω γὰρ ἀναγκαζόμενος καὶ ἢ βία πράξει ἢ οὐ φύσει ὅταν κακὸν ἀγαθοῦ ἕνεκα ἢ μείζονος κακοῦ ἀπολύσεως πράττη, καὶ ἄκων γε. οὐ γὰρ ἐφ' αὐτῷ ταῦτα. (EE 1225a18-20)

But surely it is the simpler remarks, at the beginning of the EE account and the end of the NE account, which are more correct. Actions done under duress are done voluntarily and not, strictly speaking, under any sort of compulsion, as application of the doubly modified rule of thumb clearly shows. The ἀρχὴ of an

action done under duress is within, not outside, the agent, who consents to what he does and is not overpowered by any urge like ἐπιθυμία. Hence actions under duress are not μικταὶ but simply voluntary. As for considerations about second order choice: Aristotle's own observation (NE 1110a13) that it is the chosenness of the action at the time at which it is done seems to render these irrelevant. The question whether one would choose to be confronted with such choices is a question about the nature, not of *voluntariness*, but of εὐδαιμονία. ἡ δὲ ἀνάγκη ἢ βία παύειν. (EE 1225a23-26)

It follows that action under duress is more sharply distinct than Aristotle recognises from action under compulsion, for one is a species of involuntary, the other of voluntary, action. That this distinction is not sharp for Aristotle is evident from the fact that he does not, in the NE or EE, even have separate terms for 'compulsion' and 'duress'. In EE 2.8, Aristotle does regularly use two words for 'compulsion and duress', βία καὶ ἀνάγκη (whereas, in NE 3.1, the talk is almost all of βία, apart from ἀναγκάσαντα at 1110a28). But even when Aristotle uses both words, he uses *both for both phenomena*. So EE 1225a2 (where he switches from talking about compulsion to talking about duress): λέγονται δὲ κατ' ἄλλον τρόπον βία καὶ ἀναγκασθέντες πράξει... (cp. EE 1224a11-12). (In the *Magna Moralia*, however, it is interesting to note that this distinction *is* drawn by the author (who is not necessarily Aristotle): v. MM 1188a37-b28.)

2e. Internal Duress

I quoted EE 1225a25-33 above in discussing internal compulsion, but the keen eyed reader will have seen that the immediately preceding passage, a23-25, is not about internal compulsion at all, but internal *duress*:

καὶ μᾶλλον ἂν δόξειε βίᾳ καὶ ἄκων πράττειν ἵνα μὴ ἀλγῆ ἰσχυρῶς ἢ ἵνα μὴ ἡρέμα, καὶ ὅλως ἵνα μὴ ἀλγῆ ἢ ἵνα χαίρη. (EE 1225a23-25)

It seems clear from this passage that (in the EE at least) Aristotle would admit, under the heading of duress, not only the possibility of acting so as avoid torture *by another*, but also the possibility of acting so as to avoid torture *by* (e.g.) *one's own unsatisfied longings*. EE 1225a25-28, then, would give Aristotle's rationalisation of this δόξα: there can be such internal duress, and hence it can be admissible to act so as to avoid a worse alternative which one's own constitution presents. This will only be true in some cases, of course. Only some internal urges will genuinely leave one with no practical alternative to acting on them. Other urges may make it practically very difficult to do otherwise than act on them; yet giving into them will not be in any sense a practical necessity. Therefore such giving in will be a proper subject neither of ἔπαινος nor even of συγγνώμη (NE 1110a24). (Aristotle would hardly agree with William Blake's *Proverb of Hell* that it is 'better to murder an infant in its cradle than nurse unacted

desires'.)

(If, incidentally, a25-28 is really about internal *duress*, then it might be said that a25-28 cannot be relevant in the context of internal *compulsion* in which I quoted it above. However, given Aristotle's lack of an adequate distinction between compulsion and duress in general, and his toing and froing between the two topics at EE 1225a3-33 in particular, there is no good reason why a25-28 cannot be relevant to both cases.)

2f. Duress and Persuasion

Duress, then, is not very like compulsion. It is much more like something else to which Aristotle also alludes very briefly in the same context, persuasion:

δοκεῖ δὴ τὸ βίαιον καὶ τὸ ἀναγκαῖον ἀντικεῖσθαι, καὶ ἡ βία καὶ ἡ ἀνάγκη, τῷ ἐκούσιῳ καὶ τῷ πειθοῖ ἐπὶ τῶν πραττομένων. (EE 1224a13-16)

It might at first sight be thought from this remark that Aristotle has the following spectrum of cases in mind:

τὸ βίαιον / τὸ ἀναγκαῖον // τὸ πειθός / τὸ ἐκούσιον

There would be a spectrum from entirely voluntary action (τὸ

ἑκούσιον), through voluntary action under persuasion (τὸ πειθός) and under duress (τὸ ἀναγκαῖον), to entirely compelled 'action' (τὸ βίαιον). But clearly Aristotle is not thinking of such a spectrum. As we have already seen, he makes no distinction in sense between βία and ἀνάγκη. Given which, it seems probable that no important distinction is made at EE 1224a13-16 between τὸ πειθός and τὸ ἑκούσιον either. This also is a distinction which Aristotle should have made. Action at another's instigation through persuasion is, it seems clear, quite often distinct as to its voluntariness from action which an agent simply instigates himself. That 'I was persuaded to do as I did' can perfectly well be a partially exonerating factor of a kind.

A second reason why the above spectrum will not work as a tool of analysis is that there is no easy gradation between compulsion and duress; they are not elements of the same spectrum at all. The same, however, is not true of the other three components. For (notoriously) being persuaded can take many forms, some of which are very like choosing to act purely on one's own initiative, while others are so compelling and unfair as to be close to duress. On the one side, emotional blackmail does not seem very different from any other sort of blackmail. On the other, there is evidently a close similarity between acting on another's presentation to me of reasons for action, and acting on my own presentation to myself of reasons for action. So then persuasion shades into duress on the one side, and into spontaneous voluntary action on the other. However- *pace* Aristotle- there is no such slide from compulsion to

duress. For in the range from duress, *via* persuasion, to uninfluenced original action, it remains true that the ἀρχὴ of the agent's behaviour is within the agent himself— however strong the influences which are brought to bear on that behaviour. In the case of compulsion, however, the ἀρχὴ of the agent's behaviour is (by definition) *outside* the agent; the agent is, at best, only an *influence* upon that behaviour and *not* its ἀρχὴ.

Then if duress is so different from compulsion, should Aristotle have discussed it at all in these contexts? The answer to that is, I believe, Yes. But this is not because action under duress is a species of action under compulsion; nor yet because action under duress is a species of *the involuntary*. As I hope I have shown, action under duress is neither of these. It is because action under duress is a species of *action having limited responsibility*. What the existence of the class 'actions done under duress' illustrates, rightly understood, is that an action can be fully voluntary, yet not fully responsible. This is a remarkable and counter-intuitive fact, and perhaps it was discomfort with this idea which caused Aristotle to try to fit duress under the heading of compulsion. Under that heading, however, it gets more and more anomalous the more one looks at it.

But if it is true that action under duress is voluntary, yet has limited responsibility, does it follow that *all* duress exonerates? Apparently not: in both NE and EE Aristotle notes that, just as there can be degrees of forceful influence on my

behaviour which yet do not amount to compulsion because they are not irresistible, so there can be degrees of duress which are not sufficient to justify the action performed under that duress. So at EE 1225a14-16, Aristotle remarks that it would be absurd (γελοῖος) for a person to say that he had killed someone under duress because, had he not done so, he would have lost a game of blind man's buff: ἀλλὰ δεῖ μείζον κακὸν καὶ λυπηρότερον εἶναι δπεῖσεται μὴ ποιήσας (a17). And again at NE 1110a27-29, τὸν Εὐριπίδου Ἀλκμαίωνα γελοῖα φαίνεται τὰ ἀναγκάσαντα μητροκτονῆσαι. In general, 'It is the mark of a small spirit to endure what is shameful for the sake of nothing noble or proportionate' (NE 1110a23). If one is to go through some painful or despicable course to some desired end, the end must be *worth* the pain or shame of the course. In cases where it is worth it, there can even be praise for the person who takes the necessary course (NE 1110a20). But some courses are so shameful that *nothing at all* could make them worth taking (NE 1110a26-27)-Alcmaeon's action in Euripides being a case in point. All of this evidence tends to reinforce my suggestion that action under duress is indeed voluntary action.

3. The Varieties of Ignorance

τὸ δὲ δι' ἄγνοιαν οὐχ ἑκούσιον μὲν ἅπαν ἐστίν (NE 1110b18). Aristotle's method in discussing ignorance and the voluntary too is to begin with a very plausible general rule, and then note a whole list of exceptions to it.

The plausible generalisation here is: What is done in ignorance cannot be voluntary. But this bald observation is patient of much refinement and qualification, for it is set (as the last quotation is set) in the midst of an impressive variety of both broad and subtle distinctions.

What tends to emerge from the drawing of these distinctions is that knowledge and ignorance are not set in binary opposition for Aristotle. There is a wide range of *degrees* of knowledge/ignorance, set between extremes which Aristotle (to an extent) defines for us. Hence, insofar as they depend upon knowledge/ignorance, neither are voluntariness/ involuntariness in binary opposition.

The first distinction which, strictly speaking, we should make about ignorance is one which Aristotle does not make. Aristotle uses the word *ἄγνοια* for 'ignorance'; but, as Kenny rightly notes, 'It is clear that *ἄγνοια* includes not just lack of knowledge, but also positively mistaken belief' (*Aristotle's Theory of the Will*, p.48). *ἄγνοια* covers both <believing that not-p when p is true> and <not believing that p when p is true>. It is *ἄγνοια* if (say) I switch off the lights on the mistaken belief that there is no one else in the room. It is also *ἄγνοια* if I simply switch off the lights without considering whether there is anyone else there.

The first distinction which Aristotle makes excludes a large class of cases of ignorance, as being of a kind which is *not*

relevant to the (in)voluntariness of any behaviour. This is Aristotle's distinction between what I shall call *principle ignorance* and *particular ignorance*.

3a. Principle and Particular Ignorance

This distinction is most clearly drawn at NE 1110b27 ff.. Aristotle admits the Platonist view that it is proper to speak *cognitively* of moral beliefs. We can indeed say that

ἀγνοεῖ μὲν οὖν πᾶς ὁ μοχθηρὸς ἃ δεῖ πράττειν καὶ ὧν ἀφεκτέον, καὶ διὰ τὴν τοιαύτην ἀμαρτίαν ἄδικοι καὶ ὅλως κακοὶ γίνονται. (NE 1110b27-29)

However, the Platonist is wrong, in *this* case, to draw the normally permissible inference from the μοχθηρὸς' -being ignorant (to some degree) to his being involuntary (to the same degree). τὸ δ' ἀκούσιον βούλεται λέγεσθαι οὐκ εἴ τις ἀγνοεῖ τὰ συμφέροντα (NE 1110b30). Ignorance of τὰ συμφέροντα, or 'ignorance in the proairesis' (ἡ ἐν τῇ προαιρέσει ἀγνοια, b31), is not the cause of *involuntariness*, but of *wickedness* (οὐ γὰρ αἰτία τοῦ ἀκουσίου, ἀλλὰ τῆς μοχθηρίας, b32). And Aristotle immediately makes it clear that by ἡ ἐν τῇ προαιρέσει ἀγνοια he means precisely ignorance of principle as opposed to ignorance of particulars:

...οὐδ' ἡ καθόλου [ἀγνοια αἰτία τοῦ ἀκουσίου] (ψέγονται γὰρ διὰ γε ταύτην), ἀλλ' ἡ καθ' ἕκαστα, ἐν οἷς καὶ περὶ ἃ ἡ πράξις... (NE 1110b34-1111a1)

Principle ignorance, then, ἡ καθόλου ἄγνοια, is mistaken or absent belief about what is good. Does this imply that Aristotle has any hard and fast prescriptive rules to offer us about what is good— any universal moral rules? I believe that the answer to that is No: Aristotle's moral principles are more like virtues than rules (NE 1178a18). But the argument for that conclusion will have to wait until Ch.2.

What about particular ignorance, ἡ καθ' ἕκαστα ἄγνοια? At NE 1111a2-19 Aristotle distinguishes six respects of particular ignorance:

- (a) τίς (ignorance of the *agent*),
- (b) τί (ignorance of the *act*),
- (c) περὶ τί ἢ ἐν τίνι πράττει (ignorance of the *scope* of the act),
- (d) τίνι, οἷον ὀργάνῳ (ignorance of the *means* for the act),
- (e) ἕνεκα τίνος, οἷον σωτηρίας (ignorance of the *result* of the act),
- (f) πῶς, οἷον ἡρέμα ἢ σφόδρα (ignorance of the *manner* of the act).

I think it would be implausible to suggest that this was meant as an exhaustive and definitive list, in spite of NE 1111a3 τίνα καὶ πόσα ἐστί. It might be doubted whether any list of respects in which one might possibly experience particular ignorance could be either. (Compare EE 1225b1-8, where Aristotle distinguishes only δὲν (= c), ᾧ (= d), and οὐ ἕνεκα (= e).)

Aristotle's purpose is simply to give examples of the kind of accidental unawareness of specific circumstances which counts as particular ignorance.

For consider the objection to this account that we could redescribe most cases of particular ignorance as exemplifying any one (or all) of Aristotle's respects of ignorance. Thus fencing with a sharp foil under the mistaken impression that it has a button on the end fits most obviously into Aristotle's respect (d). But we can also call it (a) ignorance that I am an agent with a sharp foil, (b) ignorance that my act is fencing with a sharp foil, (c) ignorance that the victim of my thrust is being stabbed, not touched, (e) ignorance that my act will result in death, not exercise, and (f) ignorance that I am fencing too hard.

Hence it seems that action in ignorance is more to do with mistakes of *one* type—namely, about the extension of terms—than with *six* types of ignorance. But this is not a very important objection. Aristotle's main point is not (implausibly) that there are rigidly separable *kinds* of problems about the extensional knowledge relevant to any behaviour, but simply that there may at one time be a great *variety* of different problems (of whatever kind) about that extensional knowledge. What matters is not the number of (somehow?) individuable *respects* in which the agent can truly be described as ignorant, but the number of *relevant true propositions* of which that agent is ignorant.

Two questions arise about this distinction between particular and principle ignorance. First, is it genuinely Aristotle's? Second, why should we say, with Aristotle and against some sorts of Platonist, that blame attaches to all actions on principle ignorance, but only to some actions on particular ignorance?

On the first question: Amélie Rorty has argued that Aristotle does not, in fact, have a hard distinction between principle and particular ignorance.

'Aristotle's position is far more subtle. He does not draw a sharp distinction between fact and value... For instance, the failure of knowledge in the minor premiss may be a failure of knowledge about what kind of man one is (dA 434a16-21)... but this means forgetting what is good for that type of constitution. To forget, in the face of sweets, that one is diabetic is to forget what is good for oneself.'

(Rorty 1980a, p.273)

However, (i) Rorty does not discuss NE 1110b27 ff., cited above, where Aristotle *does* (to my mind) make a sharp distinction between principle and particular ignorance, and hence a distinction (of his own sort) between fact and value. Anyway, (ii) this distinction is crucial to, and at least implicit throughout, Aristotle's whole theory of practical reasoning. Further (iii), Rorty's claim (*loc.cit.*) that 'the failure of knowledge in the minor premiss may be a failure of knowledge

about what kind of man one is' is based on the (not very decisive) evidence of dA 434a16-21. But this passage has to be weighed against Aristotle's claim (NE 1111a7) that only a $\mu\alpha\iota\nu\acute{o}\mu\epsilon\nu\omicron\varsigma$ could be ignorant of herself as agent. The decisive difficulty with Rorty's claim- if indeed she means it as a general claim- remains. This is that surely, on Aristotle's view, the awareness that one is a diabetic is *particular* knowledge; while the awareness of what one ought to do or not do *qua* diabetic is *principle* knowledge. Aristotle's claim would be that knowing this piece of particular knowledge *entails* (or should entail) knowing this piece of principle knowledge. It would not be that the two pieces of knowledge are *identical*.

The second question was: Why should we take Aristotle's line about ignorance of principle? His line is, apparently, that particular ignorance normally excuses, but principle ignorance never does. I can *always* be blamed for my ignorance of right principle; I always ought to have had this sort of knowledge. Does this requirement seem unreasonable? Then consider these two exchanges:

(1) A. You shot my father!

B. Oh, sorry- is killing wrong, then?

(2) A. You stole my watch!

B. Oh, sorry- I forgot that stealing is wrong!

Are either of B's responses conceivable as *excuses*? We can

understand B excusing himself with 'But I didn't mean to', or 'I didn't know it was your watch' (etc.); but how could 1B and 2B ever count as excuses, rather than as lunacy or as frank admission of guilt?

There is one analogy for knowledge of moral principle which Aristotle clearly has in mind at NE 1142a27 ff., and also at 1147a21, a27; though, as we shall see in Ch.2, §2, this analogy is heavily qualified by Aristotle, e.g. at NE 1140b23-24. This is the analogy between knowledge of moral principle and mathematical knowledge. No one calls me hopeless at mathematics if I don't happen to know what Goldbach's conjecture is; but they *do* call me hopeless at mathematics if I am incapable of *learning* what Goldbach's conjecture is. I may lack all sorts of factual mathematical knowledge, yet still be a genius at applying the principles of mathematics. But if I have lots of factual knowledge about maths, but no understanding of mathematical principle, then there is something radically wrong with me as a mathematician. Likewise, in ethical matters, to lack particular knowledge may affect my voluntariness, but will not affect my virtue. Whereas to lack principle knowledge will not affect my voluntariness, but will affect my virtue- for I will, *ipso facto*, be wicked. (Hence all the remaining distinctions which Aristotle makes apply only to particular and not to principle ignorance.)

In short, there is a strong contrast in Aristotle between mere knowledge *of* facts- whether learnt by definition or by perception (see Ch.2)- and knowledge *of what to do with* facts.

And this brings us to the next distinction drawn by Aristotle.

3b. Active Ignorance and Passive Knowledge

This distinction is apparent at EE 1225b11-12:

...ἐπεὶ δὲ τὸ ἐπίστασθαι καὶ τὸ εἰδέναι διττόν, ἐν μὲν τὸ ἔχειν, ἐν δὲ τὸ χρῆσθαι τῇ ἐπιστήμῃ...

Aristotle argues here (a point of some importance, as we shall see in Ch.4, for Aristotle's discussion of *akrasia*) that when some item of knowledge is only potential and not actual, it is all right to say that the agent acts in ignorance of it. This is surely true. Relevant knowledge which is not accessible to me at the time of action- even if it is then 'there in my head' in some potential form- is not properly speaking knowledge which I *have*. The agent who 'has' merely passive knowledge of a relevant particular is, then, exculpated- unless, says Aristotle, the agent *could have* actualised it (EE 1225b16-17). Which brings us to

3c. Action *on* and *in* Ignorance.

A third distinction about ignorance which Aristotle makes (NE 1110b25-30) is between τὸ δι' ἄγνοιαν πράττειν ('acting *in* ignorance') and τὸ ἀγνοῶν πράττειν ('acting *on* ignorance'):

ὁ γὰρ μεθύων ἢ ὀργιζόμενος οὐ δοκεῖ δι' ἄγνοιαν πράττειν ἀλλὰ διὰ τι τῶν εἰρημένων, οὐκ εἰδῶς δὲ, ἀλλ' ἄγνοῶν. (NE 1110b26-27).

The point of difference here, which my translation is designed to bring out, is this. An ordinarily ignorant agent (ὁ δι' ἄγνοιαν πράξας, b20), acts *in* ignorance. When he acts, he is unaware of some fact or other which is relevant to his action. Since he (of course) does not know that he is so ignorant, *a fortiori* he never chose to be so ignorant, and so deserves ἔλεος καὶ συγγνώμη (NE 1111a2). A drunkard or a man in a fury, in contrast, acts *on* ignorance. Not only does he act under influences which are liable to blind him to relevant facts; normally, he also *knows*, when so acting, that he is acting under influences which are liable to blind him to relevant facts; or at any rate he knew that it was going to be so when he began to allow himself to become drunk or enraged.

Hence, at least to a degree, ὁ ἄγνοῶν πράττων has *chosen* his state of ignorance. He has allowed himself to get in a state where there is a real danger that his actions will be importantly compromised by ignorance. And such voluntary ignorance or risking of ignorance is, according to Aristotle, culpable (*because*— note well— it is voluntary):

ὁμοίως δὲ καὶ μὴ ἔχων τις [τὴν ἐπιστήμην] ψέγοιτο ἄν, εἰ δὲ ῥάδιον ἢ ἀναγκαῖον ἦν μὴ ἔχει δι' ἀμέλειαν ἢ ἡδονὴν ἢ λύπην. (EE 1225b14-17; cp. NE 1114a2-3)

3d. Repentance and Disowning

The fourth and last of Aristotle's distinctions about under what circumstances an action in ignorance may be excused, applies only to actions done *in* ignorance, and in *particular* ignorance:

τοῦ δὴ κατὰ τὴν τοιαύτην ἄγνοιαν ἀκούσιου λεγομένου ἔτι δεῖ τὴν πράξιν λυπηρὰν εἶναι καὶ ἐν μεταμελείᾳ. (NE 1111a21)

If an action done δι' ἄγνοιαν is not followed by λυπή and μεταμελεία, then it may be οὐχ ἑκούσιον, *not voluntary*, but it is not ἀκούσιον, *involuntary* (1110b18–25). This use of μεταμελεία does not refer to an *emotion* of repentance or regret; it marks a logical distinction resting on a counterfactual. If A does x in ignorance of p, then

A does x ἀκούσιως if A would not have done x if A had known p;

but A does x οὐχ ἑκούσιως if A would still have done x even if A had known p.

For practical purposes, therefore, the question is: will A *disown* the deed x? If A will disown x, then A did x fully involuntarily, and so, *ceteris paribus*, is exculpated. But if A refuses to disown x, then A is not only not exculpated; A is actually incriminated to a degree. For the fact that A will not disown the deed suggests something about A's character. It

suggests that A is, in general, the kind of person to do this sort of thing, even though A's doing of x was not (on *this* occasion) a voluntary action. Hence, where A shows no μεταμελεία for having done x δι' ἄγνοιαν, A is not as fully culpable for x as if A had actually meant to do x. But neither is A as fully exculpated from x as A would be if A's doing of x had been involuntary rather than non-voluntary. Aristotle's emphasis on the development of character shows here. If x is blameworthy, then to be the kind of person who *would have done x* is itself to be a proper subject of a degree of blame- but not of the same degree of blame as the person who actually *does x*. Aristotle marks the difference in degree of voluntariness between τὸ οὐχ ἔκούσιον and τὸ ἀκούσιον precisely because he wants to mark a corresponding difference in degree of responsibility. (John Lucas has argued (in unpublished lectures on 'Responsibility', given at Oxford in Trinity Term 1991 and 1992) that what Aristotle was expounding at NE 1110b18-24 was a doctrine of 'retrospective culpability' seen as a case of 'Cambridge change'. I can agree provided it is not assumed that τὸ οὐχ ἔκούσιον is *just as culpable as* τὸ ἀκούσιον.)

Aristotle's decision procedure for determining whether an agent who acts in some kind of ignorance is or is not a voluntary agent, and so is or is not to be exculpated, can be summarised in this flow chart:

- A. Is the ignorance of principle(s) or of particular(s)?
 - A1. Of principle(s): agent acts voluntarily and hence culpably.

A2. Of particular(s): proceed to B.

B. Is the apparent 'particular ignorance' in fact active ignorance which the agent could have helped? Or is it a case of passive knowledge? Or neither?

B1. Active ignorance: agent acts voluntarily and hence culpably.

B2. Passive knowledge: proceed to D.

B3. Neither: proceed to C.

C. Is the action action *in* particular ignorance or *on* particular ignorance?

C1. On: agent acts voluntarily and hence culpably.

C2. In: proceed to D.

D. Does the agent show μεταμελεία?

D1. No: agent acts semi-voluntarily and hence semi-culpably.

D2. Yes: agent is exculpated.

This concludes my exegesis of Aristotle's treatment of the relation of free action to (i) compulsion and (ii) ignorance. By now certain broad themes of this thesis are emerging. One is that, although the desire to ascribe praise and blame is not (on my view of Aristotle) the only possible motivation for an inquiry into what factors make for (in)voluntariness, it is evident that, for him, voluntariness and responsibility stand in direct proportion. Another is that there can be degrees to ignorance, compulsion and duress, and hence also to (in)voluntariness and responsibility.

4. Irrationality

I come to the third negative condition of voluntary action which, I said, the logic of Aristotle's position requires. Saying that 'the logic of Aristotle's position requires' this third negative condition means, of course, that I am going to argue for it myself, because Aristotle does not, at least not directly. But that there is here a hole which Aristotle should have filled is not too difficult to show.

Compare his negative conditions of voluntary action with his positive account. In this (exegesis of which is yet to come, in Chs.2-3), Aristotle will tell us that voluntary action is

(i) action which originates within the agent in a particular way (a condition corresponding to the negative condition about compulsion);

(ii) action on certain sorts of knowledge and information (a condition corresponding to the negative condition about ignorance);

and

(iii) action which is *rational behaviour*.

The negative condition to which this third part of the positive account corresponds is the condition that voluntary action must not be *irrational behaviour*, a phrase the sense of which I now hope to make clearer. The argument for positing this third negative condition spills over unavoidably into the

forthcoming positive account; but here it is, prolepses and all.

To put the point in a slogan: Belief *plus* want does not equal voluntary action; but belief *in combination with* want does. On Aristotle's own theory, there is in the constitution of a voluntary action more than just the absence of ignorance (which implies, positively, the presence of true belief(s) relevant to the manifested behaviour) and the absence of compulsion (which implies, positively, the presence of unhindered desire(s) relevant to the manifested behaviour). It is neither sufficient nor necessary for the occurrence of a voluntary action that the agent's true belief(s) and unhindered desire(s) should *be* relevant to the performance of the manifested behaviour. What is sufficient and necessary for a voluntary action is that those belief(s) and desire(s) should *be seen by the agent herself to be* relevant to the performance of the manifested behaviour.

Consider a person A who exhibits some behaviour B. Let us suppose (i) that there is nothing in the situation to impede B. Let us also say (ii) that, in every important respect relevant to B, A is not ignorant. Does it follow, from the stipulation of these two conditions alone, that B is a *voluntary action*? If it does follow, then my negative conditions (i) and (ii) are both necessary and sufficient for voluntary action. But if it does not follow, then (i) and (ii) may be necessary, but they are not necessary-and-sufficient for voluntary action.

My argument for negative condition (iii) is that it does *not*

follow. For if A, in the situation described, is to do B as a *voluntary action*, this condition is not necessary-and-sufficient:

(C1) that A should do B while feeling unimpeded relevant desire(s), and while holding some relevant belief(s) not subject to any kind of ignorance.

What is necessary-and-sufficient for A's performance of B as a voluntary action is *this*:

(C2) that A should do B because of the combination of some unimpeded desire(s) which A feels, and some relevant belief(s), not subject to any kind of ignorance, which A holds.

As my stresses indicate, the crucial difference between (C1) and (C2) is in the words 'because of the combination of'. To say that A does B *because of the combination of* desire D and belief P is to say something quite different from, and not implicit in, saying that A does B *while feeling* desire D and *while holding* belief P. All my negative conditions of voluntary action (i) and (ii) entitle us to say is the latter and not the former.

This is why, on their own, (i) and (ii) *cannot* be the necessary and sufficient conditions of voluntary action. For the satisfaction of (i) and (ii) by some piece of behaviour is no guarantee that that piece of behaviour can plausibly be seen as a voluntary action rather than as *just* a piece of (random) behaviour. I might want some toast, and believe rightly that this

thing here is toast, and go ahead to eat the toast without being compelled to do so. Yet, in this case as described, it could still be true that I did not eat the toast *because* I wanted toast and believed this thing to be toast. There might be no connection at all between my beliefs and desires and my toast-eating behaviour. That behaviour might be purely coincidental to those beliefs and desires.

Of course (it may be replied), someone for whom there was no such connection would be counted, by most people, as an idiot or lunatic. But that is precisely the point: that negative conditions (i) and (ii), taken on their own, give us *no means whatever* of distinguishing the chance, random movements of a lunatic (or certain sorts of machine) from the voluntary actions of a rational agent. To make that distinction, we need to postulate a third negative condition of voluntary action: that it should not be irrational behaviour.

Note that behaviour which is irrational, in this sense, is *not* simply a sub-class of behaviour which is compelled— although that seems to be Aristotle's implication when he writes of the ἐνθουσιῶντες that, even though their behaviour is actually intelligent (καίπερ διανοίας ἔργον ποιοῦντας, EE 1225a29), 'still we do not say that it is up to them to say what they say or do what they do'. Likewise, Aristotle describes incontinent people as being like actors (τοὺς ὑποκρινομένους, NE 1147a23), and as being like the sleepwalker or the madman or the drunkard (μαινόμενος, καθεύδων, οἰνωμένον, NE 1147a14). Now he says



(rightly, I think) that the behaviour of people in such states is not fully voluntary; but he seems to be implying (wrongly, I suggest) that such behaviour is not fully voluntary *because it is compelled*. In the case at least of the sleepwalker and the madman, it would have been much more plausible for him to say rather that such behaviour is not fully voluntary *because it is irrational*. In such cases, that is, the connections which we would look for, in a voluntary action, between wants and beliefs and the manifested behaviour are simply absent.

For irrational behaviour differs from compelled behaviour (as that is described in Aristotle's own account) for the following reasons. Compelled behaviour, we have seen, may (a) originate outside the agent. But irrational behaviour can perfectly well originate within the agent, and still be irrational; hence irrational behaviour is not externally compelled. Otherwise it may (b) originate within the agent, against the overpowered agent's consent. But that formula implies that something like a normal process of combining wants and beliefs towards an action goes on in the agent, but is frustrated by an overriding force—otherwise there would be no consent for the compelled behaviour to be against. If no such process occurs, as (by definition) it does not in irrational behaviour, then there is no such consent for the resulting behaviour to override; hence irrational behaviour is not internally compelled either.

Nor is irrational behaviour a sub-class of action in ignorance. *That* might seem to be Aristotle's implication when he

describes the person who fears nothing as a *μαινόμενος* (NE 1115b27). Fearlessness might be thought to count as a case of ignorance-like irrationality in the context of Socrates' famous argument (*Laches* 194c-d) that courage was a species of knowledge. Again, Aristotle calls a *μαινόμενος* the person who is ignorant of every aspect of his own action (NE 1111a7).

Now no doubt crazy courage, and such radical ignorance as Aristotle means, are cases which do deserve the name *μανία*. But that does not mean that ignorance of either of these sorts is equivalent to what I mean by 'irrationality'. In a case of irrational behaviour in my sense, the problem is not what it is in a case of ignorance (as that is described in Aristotle's own account). It is not that the agent has no beliefs, nor that the agent has wrong beliefs. The problem is that the agent's beliefs fail to connect in any conceivably appropriate way with the agent's behaviour.

It is because irrational behaviour will not fit under either of *Aristotle's own* headings of ignorance or of compulsion that I say he needs to postulate a third negative condition of voluntary action. Does Aristotle himself ever recognise the need for that third negative condition?

The answer to that seems to be: hardly at all. It is well known that Aristotle has a theory of practical reason (to which I shall be coming shortly) in which a certain form of voluntary action, action on what is called *προαίρεσις* by Aristotle, plays a

starring role. Action on προαίρεσις is, evidently, voluntary action which is rational; but it is also sharply distinguished by Aristotle from the voluntary in general (τὸ ἐκούσιον). Moreover, at first sight it looks as if the point of distinction between τὸ ἐκούσιον and action on προαίρεσις is exactly that action on προαίρεσις is necessarily rational, while τὸ ἐκούσιον is *not*. (So, e.g., NE 1112a13-17.)

But my claim is not about action on προαίρεσις; it is about τὸ ἐκούσιον. The claim I have just argued for is that, contrary to the apparent sense of passages like NE 1112a13-17, τὸ ἐκούσιον too is necessarily rational (in the sense defined above). (What, then, is the *right* way to distinguish between τὸ ἐκούσιον and action on προαίρεσις? I will answer that question in Ch.3, §2.) For the explicit making of *this* claim there is little or no evidence in Aristotle's ethical writings. The best evidence is to be found in the MM, but this is of dubious authorship. At MM 1188b25-28, the author does indeed give *three* conditions of voluntariness, not two:

τὸ γὰρ ἀκούσιόν ἐστι τό τε κατ' ἀνάγκην καὶ κατὰ βίαν γιγνόμενον, καὶ τρίτον δὲ μὴ μετὰ διανοίας γίγνεται.

But clearly the three conditions of a voluntary action given here are not my three. They are the conditions that the action should involve (a) no compulsion, (b) no duress, and (c) the presence of 'understanding': meaning what? διανοία might easily be the Greek for 'rationality' (in a loose sense); but it becomes

clear, when we read MM 1188b29-35, that what the author means by requiring the presence of διανοία in a voluntary agent is very little different from what Aristotle means by requiring that a voluntary agent should be εἰδώς.

The only places in the writings generally accepted as genuinely Aristotle's to which one might, tenuously, point for direct evidence of my third condition would be these two:

Λεκτέον δ' ἴσως βουλευτὸν οὐχ ὑπὲρ οὗ βουλευσάιτ' ἂν τις ἡλίθιος ἢ μαινόμενος, ἀλλ' ὑπὲρ ὧν ὁ νοῦν ἔχων. (NE 1112a20)

[τὰ θηρία] οὐ γὰρ ἔχει προαίρεσιν οὔδε λογισμόν, ἀλλ' ἐξέστηκε τῆς φύσεως, ὥσπερ οἱ μαινόμενοι τῶν ἀνθρώπων. (NE 1149b35)

The first passage might be adduced to support the claim that it is only the μαινόμενος, meaning the irrational agent in my sense of 'irrational', who does not reason in his voluntary action. But this will not do. (i) The passage says rather that the μαινόμενος *does* reason, but not about the right things; clearly then this kind of μαινόμενος cannot be the 'irrational agent' of my terminology. (ii) The passage is taken from the beginning of Aristotle's discussion of βουλή, which is for him an ingredient of προαίρεσις (NE 1112a15).

The second passage might be taken to mean that what the animals have in common with the μαινόμενος (again, meaning the

irrational agent in my sense of 'irrational') is that neither acts voluntarily *because* neither reasons: the animals because they act on instinct or the like, the μαινόμενος because he is a μαινόμενος. But this will not do either. (i) The passage, explicitly, is talking about exceptions from προαίρεσις, not from τὸ ἔκούσιον. (ii) Aristotle says at NE 1111b8-10 that animals, like children, *do* share in the voluntary, but not in action on proairesis: τοῦ μὲν γὰρ ἔκουσίου καὶ παῖδες καὶ τᾶλλα ζῷα κοινωνεῖ, προαιρέσεως δ' οὔ.

So neither of these passages can be used to give direct support to the claim, which I nonetheless have argued for, that Aristotle is or should be committed to the doctrine that *all* voluntary action is rational (in my sense of that word). (This second passage, moreover, returns my analysis to the question raised above: What, on my reading, is the *right* way to distinguish between τὸ ἔκούσιον and action on προαίρεσις? V. Ch.3.) However, this absence of 'direct' support, such as would be given by the discovery of an Aristotelian term meaning 'rational' in my sense, is no problem. As Ch.3 will show, my claim is not short of other, less immediately obvious forms of grounding.

Let me conclude this chapter by noting that, in explanation of my third negative condition of the voluntary, I can now essay a tentative definition of rational and irrational behaviour. The definition is this: An agent's rational behaviour is such of that agent's behaviour as is susceptible to explanation by reference

to the efficient causal influence of the combination of that agent's wants and beliefs, where that combination has an *efficient causal* influence over the *behaviour* because it *provides a reason* for the *action*.

Irrational behaviour, then, is just behaviour which cannot be explained in the above way: human movement or doing for which there is *no* contrast between two ways of describing it, as behaviour, with its efficient causes (on the one side), and as action, with the reasons for which it is done (on the other). And if behaviour which satisfies the other two conditions of voluntary action *is*, nonetheless, irrational behaviour, there seems, as we have seen, to be no good reason to describe it as *voluntary* behaviour rather than (say) random, accidental, instinctive, or automatic behaviour. This doctrine of the necessary *rationality* of voluntary action is one for which I would claim the support of Donald Davidson. In his words:

'Central to the relation between a reason and the action it explains is the idea that the agent performed the action *because* he had the reason... When we ask why someone acted as he did, we want to be provided with an interpretation. His behaviour seems strange, alien, outré, pointless, out of character; *perhaps we cannot even recognise an action in it*. When we learn his reason, we have an interpretation, a new description of what he did, which fits it into a familiar picture [which]... includes some of the agent's beliefs and attitudes... To learn that the agent conceived his action as a lie, a repayment of a debt, an insult,

the fulfilment of an avuncular obligation, or a knight's gambit is to grasp the point of the action in its setting of rules, practices, expectations, and conventions... *And there is no denying that... when we explain an action, by giving the reason, we do redescribe the action.*'

(Donald Davidson, 'Actions, Reasons and Causes', in Davidson 1980, pp.9-10; emphases added)

Chapter 2

Freedom, Ability and Knowledge

1. Function, Process, and 'Ability to do Otherwise'
2. Aristotle's Epistemology

'It seems clear, at the outset, that Aristotle analysed voluntary action in terms of efficient causality *and* knowledge... These two elements are central to Aristotle's account: neither causation nor knowledge alone is sufficient to analyse the concept; both causal and teleological considerations may play a role in it without inconsistency. Most recent work [e.g. Anscombe (1957), von Wright (1971), Stoutland (1970, 1976)] has sought to analyse voluntary action in terms either of causation *or* of knowledge and teleology (but not both). Aristotle's theory (if it proves defensible) may avoid those difficulties which undermine contemporary accounts which take one of these features alone as the analysans; for it would yield a (partially) causal account of intentional action and the basis for a (partially) causal analysis of freedom to act.'

(Charles 1984, p.59)

In the last chapter, in my account of the three negative conditions of voluntary action to which Aristotle is committed, I

remarked that each of these conditions has a corollary which is a part of Aristotle's positive account of voluntary action. In this chapter, I begin my exegesis of that positive account by examining the positive corollaries of these negative conditions about (i) compulsion and (ii) ignorance. Treatment of the corollaries of (iii), the condition about irrationality, will come in Ch.3.

1. Function, Process, and 'Ability to do Otherwise'

If any human behaviour, to count as voluntary action, must be uncompelled, what does Aristotle think will count as such uncompelled behaviour? We have already glanced at his brief answer to this at EE 1224a16-20. The opposite of compelled behaviour, for any item, is whatever behaviour is *natural* for that item. Thus an uncompelled stone falls towards the centre of the earth because there is nothing in the situation to prevent it doing what it is *natural* for stones to do; and a flame, under normal conditions, rises away from the centre of the earth for the same sort of reason. Now as for flames and stones, so for humans: the way to understand what Aristotle thinks counts as uncompelled human behaviour is to ask:- What does he think counts as *natural* human behaviour?

To answer this question properly, we must look at Aristotle's theory of human nature. This theory leads Aristotle to a detailed answer to that question. It also offers us an understanding of

the place of humanity within the cosmos, and underlies a good deal of what Aristotle has to say, both about political or social man (the subject of the ethical discourses: NE 1094b7-10, 1097b8-11), and also about women, children, barbarians, animals, plants, and other genera even further removed in type from the Aristotelian ideal. It is only in the context of that wider view that the specific answer is given. We ought then to start with the wider view.

1a. Function

The starting-point for understanding Aristotle's theory of human nature is the concept of the ἔργον or function which he took over from Plato (*Republic* 352e-353a). Aristotle's doctrine of function is given at NE 1097b22-1098a20. Here Aristotle is seeking a more precise answer than 'Happiness' to the question 'What is the human good?'. He suggests that τάχα δὴ γένοιτ' ἄν τοῦτ', εἰ ληφθεῖη τὸ ἔργον τοῦ ἀνθρώπου (1097b25). The idea is that, if we look at what humanity *characteristically* and *by nature* does or is, then this will give us a clearer view of what it might be good for humanity to do:

ὥσπερ γὰρ αὐλητῆι καὶ ἀγαματοποιῶ καὶ παντὶ τεχνίτῃ, καὶ ὅλως ὧν ἐστὶν ἔργον καὶ πράξις, ἐν τῷ ἔργῳ δοκεῖ τὰγαθὸν εἶναι καὶ τὸ εὖ, οὕτω δόξειεν ἄν καὶ ἀνθρώπῳ, εἴπερ ἔστι τι ἔργον αὐτοῦ. (NE 1097b26-28)

It is an obvious feature of human life that it includes many different particular kinds of τεχναὶ, skills, in which a person might display ἀρετή, excellence. Now we can provide a simple and non-circular analysis of the meaning of 'excellence' for each of these particular skills. For excellence at (e.g.) harp-playing is indeed, in Aristotle's phrase, ἐν τῷ ἔργῳ, 'implicit in the activity' as we might translate it (note the useful ambiguities of the Greek word ἔργον). That is: the knowledge of what counts as *excellence in harp-playing* is directly deducible from the knowledge of what counts as *harp-playing, tout court*: τὸ δ' αὐτὸ φάμεν ἔργον εἶναι τῷ γένει τοῦδε καὶ τοῦδε σπουδαίου (NE 1098a8-9). To define what performing a function (ἔργον) is, just is to define what performing that function *well* is. Aristotle wants to suggest that human life itself is a function- or, to use a related Aristotelian term, an activity (ἐνέργεια)- of this sort. Defining what that activity is will enable us to deduce what doing that activity *well* is. And this will be the human good.

Aristotle's claim that humans, *qua* humans, must have an ἔργον, is here supported by two distinct arguments *a fortiori*: (i) Humans have ἔργα (even) qua carpenters, cobblers or whatever; so how can we suppose that Humanity Itself doesn't have an ἔργον? (The examples are deliberately chosen for their banal lowliness: cp. Plato, *Republic* 434b-d.) (ii) Parts of humans have ἔργα; so how can we suppose that Humanity Itself doesn't have an ἔργον?

These two arguments connect clearly enough with a much more general claim of Aristotle's, that the whole of nature displays *purpose*: τὸ γὰρ μὴ τυχόντως ἀλλ' ἔνεκά τινος ἐν τοῖς τῆς φύσεως ἔργοις ἐστὶ καὶ μάλιστα (dPA 645a24-5); cp. dGA 744a16, which Ross (1925) translates as 'Nature, like a good householder, throws away nothing of which anything useful may be made'. Aristotle also believes that the whole of nature displays *economy of design*. At PoI 1252b1-5, he gives us the 'Delphic knife' principle, that every distinct species made by nature has a single ἔργον:

οὐθεν γὰρ ἡ φύσις ποιεῖ τοιοῦτον οἷον χαλκοτύποι τὴν Δελφικὴν μάχαιραν πενιχρῶς, ἀλλ' ἐν πρὸς ἐν· οὕτω γὰρ ἂν ἀποτελοῖτο κάλλιστα τῶν ὀργάνων ἕκαστον, μὴ πολλοῖς ἔργοις ἀλλ' ἐνὶ δουλεῦον. (PoI 1252b1-5)

From this argument to the conclusion that humans, *qua* humans, have an ἔργον, Aristotle turns to the question of what that ἔργον could be. In line with his general theory of classification by *genus* and *species*, Aristotle tells us (NE 1097b35) that what is sought when one is looking for some thing's function is τὸ ἴδιον, what is *characteristic* of that thing or unique to it. Thus the human function cannot be given by the mere fact that humans are alive (τὸ ζῆν, b35), nor can it be the life of nourishing and the life of growth (τὴν τε θρεπτικὴν καὶ τὴν αὐξητικὴν ζωὴν, 1098a1). For these we have in common (κοινὸν, b35) even with plants. Nor, for similar reasons, can the human function be 'some kind of life of sensation' (αἰσθητικὴ τις, 1098a2-3); for this also is shared,

'with horses and cows and every kind of animal'. Hence it follows, or at least Aristotle thinks it follows, that λείπεται δὴ πρακτικὴ τις τοῦ λόγον ἔχοντος (1098a4; cp. dA 413a21-b11). What humans have which they share with no other creature is the ability to partake of a life of rational action. Thus, in the NE and the *Politics* at least, the way to find the ἔργον of x is to look for the *one thing* which that x characteristically does or is that differentiates x from y and z. But there are at least two problems with this.

Firstly, an exegetical problem: at dPA 642b5-643b26, Aristotle argues forcefully *against* the taxonomical method of dichotomy by a single differentia (he may, or may not, have in mind Plato's demonstration of taxonomical method at the start of the *Sophist* (219a ff.)). *Inter alia*, Aristotle objects (dPA 642b7-10) that some differences are more significant than others; and that normal, pretheoretical taxonomy κατὰ γένη is more satisfactory than any other kind (dPA 643b10-12: a typically Aristotelian verdict in favour of the ἔνδοξα; cp. NE 1098b16-18, 27-30). Thus, in the biological writings, the division of animals by species is, for Aristotle, logically prior to the assignment of an ἔργον to each species. The γένος determines the ἔργον, not *vice versa*. Moreover Aristotle presents strong arguments for not assuming that the assignment of ἔργα is a simple and mechanical procedure.

Secondly, a logical problem. In the NE, Aristotle clearly works on the assumption that, if x, y and z all φ, then φing

cannot be the ἔργον of any of them. A notorious problem follows, from this account and an examination of NE X.8, about what the ἔργον of the gods could be. Assuming that the human ἔργον is, roughly, *πρᾶξις*, the natural answer to this is apparently that *θεωρία* is the divine function. But humans are capable of *θεωρία* as well as *πρᾶξις*: so nothing is left over to be distinctive to the gods.

But the problem is wider than that. The same problem could be raised about any creature at all, other than (presumably) the human being. If ἔργον means 'what x characteristically and uniquely does or is', then plants have no ἔργον, because everything they do or are (viz., whatever is entailed by 'the life of nourishing and the life of growth') is also done by or involved in the nature of other, more sophisticated creatures. But the same will be true of, say, horses. For while they are different from plants in as much as they not only have 'the life of nourishing and the life of growth', but also 'some sort of life of sensation', they do not (according to Aristotle) have anything like 'a [life] of action of the part that has reason'. So on this understanding of ἔργον, it turns out that in fact only humans have an ἔργον; it is not just the gods who are excluded. But this hardly accords with Aristotle's doctrine of the purposefulness of all nature, mentioned above.

How are we to tidy this up? One way of producing a solution of a kind, which will not square with every piece of evidence but will make sense of a fair number of them, would be to attend more

closely to the role played in Aristotle's thinking by the concept of a *hierarchy of nature*.

1b. The Hierarchy of Nature

Aristotle never explicitly argues for this concept, except perhaps at dA 413a21-b11; he seems to feel entitled to take it for granted. But it is clear enough that the idea of a hierarchy of nature is at work in such a passage as the opening of NE VII (1145a15-33). Here Aristotle, ἄλλην ποιησαμένος ἀρχήν, suggests that 'there are three forms in respect of character which are to be avoided, wickedness, akrasia and beastliness' (θηριότης, a17). What is θηριότης, and what is its opposite? Aristotle replies: μάλιστα' ἄν ἀρμόττοι λέγειν [τὸ ἐναντίον] τὴν ὑπερ ἡμᾶς ἀρετὴν, ἥρωικὴν τινα καὶ θεῖον (a19-20), and quotes Homer, *Iliad* 24.258:

οὐδὲ ἐφύκει

ἄνδρός γε θνητοῦ πάις ἔμμεναι ἀλλὰ θεοῖο.

It seems entirely natural to Aristotle to say that the opposite disposition (ἔξις) to θηριότης is heroic or godlike virtue. Beastliness, for him, is 'below' humanity, and godlikeness is 'above' it; he even quotes the common opinion that ἐξ ἀνθρώπων γίνονται θεοὶ δι' ἀρετῆς ὑπερβολῆν (NE 1145a23), a pregnant remark to which I shall return below. The opposite is also true: one can become bestial through over-abundance of badness (a23-25). Further, Aristotle notes, barbarians (and, Pol

1252b8-9 tells us, slaves too) are more likely to be bestial than Greeks (a31-32). Finally, maimed or diseased or brain-damaged persons are less than fully human, and sometimes so much so as to be bestial (a32-33). (By this, nasty as it sounds, Aristotle need only mean that to be so crippled is to be unable to perform the human ἔργον properly, or at all; not necessarily that to be in such a state is to be a fit object of contempt.)

All this, I think, is very clear evidence that Aristotle believed in a hierarchy of nature, in which some creatures are 'higher' (and therefore superior) and others 'lower' (and therefore inferior). We can add to this evidence that of various other passages, all of which, I suggest, will continue to fill out the picture of an Aristotelian doctrine of a hierarchy of nature.

(i) Internal to the person, there are, for Aristotle just as much as for Plato, inferior and superior parts:

Pol 1254b5-7, 8: Ἡ μὲν γὰρ ψυχὴ τοῦ σώματος ἄρχει δεσποτικὴν ἀρχήν, ὁ δὲ νοῦς τῆς ὀρέξεως πολιτικὴν καὶ βασιλικήν... κατὰ φύσιν καὶ συμφέρον τὸ ἄρχεσθαι τῷ σώματι ὑπὸ τῆς ψυχῆς... (cp. Plato, *Republic* 443d-e; NE 1102a26-1103a1, 1113a5-9; VV 1249a31 ff.).

(ii) Likewise there are inferior and superior parts within the human race:

(a) Men/husbands superior to women/wives: PoI 1254b13-15: τὸ ἄρρεν πρὸς τὸ θῆλυ φύσει τὸ μὲν κρεῖττον τὸ δὲ χεῖρον, τὸ μὲν ἄρχον τὸ δ' ἄρχόμενον.

(b) Masters superior to slaves: PoI 1254b13-15.

(c) Greeks superior to barbarians: PoI 1252b8, quoting Euripides' βαρβάρων δ' Ἕλληνας ἄρχειν εἰκόσ (*Iphigeneia in Aulis* 1400); PoI 1255a28-32.

(d1) Fathers superior to children: PoI 1259b2-5, NE 1134b8-12.

(d2) Children, like animals, not full moral agents: NE 1111b8-10; cp. NE 1153a27-31.

(e) God-like, ordinary and bestial humans ranked as above.

(iii) And, lastly, there are inferior and superior parts in nature generally:

(a) Tame animals superior to wild ones: PoI 1254b10-13 (because, perhaps, they are closer to being ruled by reason; cp. NE 1102b29 ff.).

(b) Animals, like children, not (full?) moral agents: NE 1111b8-10.

For the next rungs on the ladder, we may refer to NE 1097b34-1098a4, or to dA 413a21-b11. It seems that the 'hierarchy of nature' is ordered as follows:

(c) Animals capable of all of these: νοῦς, αἴσθησις, κίνησις καὶ στάσις ἢ κατὰ τόπον, ἔτι κίνησις ἢ κατὰ τροφήν καὶ φθίσις τε καὶ αὔξεισις (dA 413a23-25); eg. (?) humans;

(d) animals capable of all of these except one: e.g. horses (NE 1098a3) are incapable of exercising νοῦς;

(e) animals capable of all of these except two: e.g. a barnacle is incapable of exercising νοῦς or κίνησις κατὰ τόπον;

(f) animals (or rather, by now, living things) capable of all of these except three: e.g. a plant has only the θεραπευτικὴν καὶ τὴν αὐξητικὴν ζωὴν (NE 1098a2; cp. dA 413a26-b10);

(g) the inanimate (superiority of the study of the soul: dA 402a1 ff.).

1c. Hierarchy and Function

How does this doctrine of the 'hierarchy of nature' help us clarify the doctrine of the ἔργον? The answer is that it suggests that we should redefine the 'ἔργον of x'. Any x's function is not simply 'what x characteristically and uniquely does or is'. It is, more exactly, 'what x characteristically and uniquely does or is *that differentiates it from what is below it in the hierarchy of nature*'.

This redefinition enables us to overcome the logical difficulty noticed above. On this definition, it will not be true that nothing except humanity has a function. Each creature in the hierarchy of nature will have a function which defines its place in that hierarchy. X, which does A and B but not C, will fit into the hierarchy between Y, which does A but not B or C, and Z, which does A, B *and* C. Here we cannot say that Z does not have A

as a function, simply because Z does A, but X does A too. Nor will it be true of X that it does not have B as a function, simply because X does A and B, but X does A and B too. The τὸ ἕδιον we are looking for when we look for a creature's function, on this account, is not simply what differentiates that creature from *any* other creature; it is what differentiates that creature from *all lower* creatures.

Of course, there are obvious problems here. (i) There may indeed be some activities x, y and z for which it is true that, say, humans perform x, y and z, but elephants only x and y (perhaps: running, eating, making hydrogen bombs). But so what? First, we could just as easily cite some activities p, q and r for which it is true that elephants perform p, q and r, but humans only p and q (perhaps: running, eating, uprooting trees with the trunk). And second, as my hydrogen bomb example suggests, the differentiating activity of the species which is capable of more activities might be an *abhorrent* activity, one which (it might be said) ought to *demote* that species in the hierarchy of nature, not *promote* it. This is a telling objection to the hierarchy doctrine, at least in any form in which the order of the hierarchy is supposed to follow from a definition of γένη in terms of ἔργα; but it does not detain Aristotle, and need not detain us.

More specifically (ii and iii), this redefinition still contradicts the 'Delphic knife' principle (Po1 1252b1-5). There are still two classes of things without any ἔργον: namely, at the

bottom of the hierarchy, whatever is left behind when all the things with ἔργα have been distinguished from it and put above it in the hierarchy; and, at the top of the hierarchy, the gods. Lastly (iv), does this procedure really tell us that the ἔργον of humanity is what Aristotle says it is?

On (ii), there is little problem about there being no ἔργον for whatever is left at the bottom of the hierarchy. This need not be a γένος, and therefore need not have an ἔργον. What is left behind by the definition and specification of all γένη and their corresponding ἔργα is simply the *undefined* or the 'unlimited'; which of course has no ἔργον, no more than τὸ ὑποκείμενον as such has any εἶδος.

On (iii) and (iv): the answers to these questions come together. Let us try (iv) first, and ask again what the human ἔργον is. Aristotle's problem, in answering this question, is not only that it is not at all clear that what he wants to call the human ἔργον has been arrived at strictly by the method for differentiating between species which he claims to be using. More troubling still, he does not even have a single answer to the question:

- (a) λείπεται δὴ πρακτικὴ τις τοῦ λόγον ἔχοντος... (NE 1098a3-4).
- (b) φύσει πολιτικὸν ὁ ἄνθρωπος. (NE 1097b11, PoI 1253a3)
- (c) ἡ τελεία δὴ εὐδαιμονία αὕτη [θεωρία] ἂν εἴη ἀνθρώπου. (NE 1177b24-25)
- (d) ἔργον [ἀνθρώπου]... ψυχῆς ἐνέργειαν καὶ πράξεις μετὰ λόγου... (NE 1098a14).

It is commonly said that Aristotle has two accounts of the good for humans, one intellectual and the other practical. But, on the face of it, the above four quotes present not just two but *four* different verdicts on the nature of the good life. (a) is clearly the practical good, (c) the intellectual, good; however, in addition to these, (b) suggests a *political* good- not necessarily the same as a practical good, as Aristotle himself admits even while trying to run them together (NE 1141b23)- and (d) suggests a *composite*, intellectual cum practical good, which of course cannot properly be identified with either of its components as represented in (a) and (c).

Is there a way of making sense of this jumble? Many commentators think not (e.g. Clark (1978) arranges the evidence in the same sort of way as I do, but jibs at the conclusions I will now draw from that arrangement). My suggestion is, first, that we suppose that some kind of valid connection can, as Aristotle believes, be made between (b) and (a)- what kind, I will not discuss here. And second, that we make sense of the relations between (a), (c) and (d) by introducing a concept which stands alongside that of the ἔργον: the concept of the *meta-ἔργον*.

What I mean to suggest by my *meta-ἔργον/ ἔργον* contrast is that any creature in the Aristotelian hierarchy of nature has, not one, but *two* final goods (τέλη). Recall that Aristotle does not just talk of excellence *within* a τέχνη, ἔργον or γένος. He also talks of the excellence *of* a τέχνη, ἔργον or γένος. That, of

course, is the whole point of the hierarchy of nature. It allows us to say *both* that some creatures, and their activities, rank higher than others, *and* that some creatures (at least at the human level) achieve the τέλος of their own rank more nearly than others.

Correspondingly, I am suggesting, the ἔργον of any creature or craft practitioner is to achieve excellence within its own rank in the hierarchy of nature. The *meta-ἔργον* of that creature or craft practitioner is *to rise to a higher rank in the hierarchy of nature.*

And there is good evidence for something like this view in Aristotle. (i) There is the point noted above, that ἐξ ἀνθρώπων γίνονται θεοὶ δι' ἀρετῆς ὑπερβολὴν (NE 1145a23).

(ii) This interpretation makes sense of the puzzling words of NE 1098a16-18: τὸ ἀνθρώπινον ἀγαθὸν ψυχῆς ἐνέργεια γίνεται κατ' ἀρετὴν, εἰ δὲ πλείους αἱ ἀρεταί, κατὰ τὴν ἀρίστην καὶ τελειοτάτην.

(iii) There is NE 1141a21-22: ἄτοπον γὰρ εἶ τις τὴν πολιτικὴν ἢ τὴν φρόνησιν σπουδαιοτάτην οἶεται εἶναι, εἰ μὴ τὸ ἀριστον τῶν ἐν τῷ κόσμῳ ἀνθρωπός ἐστιν.

(iv) Most conclusively of all, there are the famous words of NE 1177b27 ff.:

ὁ δὲ τοιοῦτος ἂν εἴη βίος κρείττων ἢ κατ' ἀνθρώπον· οὐ γὰρ ἦ ἀνθρωπός ἐστιν οὕτω βιώσεται, ἀλλ' ἦ θεϊόν τι ἐν αὐτῷ ὑπάρχει· ὅσον δὲ διαφέρει τοῦτο τοῦ συνθέτου, τοσοῦτον καὶ ἡ ἐνέργεια τῆς

κατὰ τὴν ἄλλην ἀρετὴν... ἐφ' ὅσον ἐνδέχεται χρῆ ἀθανατίζειν...
(NE 1177b27-35)

I suggest that the upshot of the argument, given the evidence of these passages, is this. Granted the coherence of the very idea of a hierarchy of nature, we may say that each γένος in it, at least in the higher reaches of the hierarchy, has both an ἔργον and a meta-ἔργον. Each γένος' ἔργον is given, as NE 1098a8-9 suggests, by the definition of that γένος, and there is a τέλος for each γένος of excellence within the ἔργον of that γένος. But there is also a meta-ἔργον (and a meta-τέλος) for each γένος, which is to *ascend the hierarchy of nature*: to rise beyond the performing of its own ἔργον to the performing of the ἔργον of the next γένος up.

The application of this line of thought to the case of human, divine and animal ἔργα should be clear. The human ἔργον, on this view, is straightforwardly not the contemplative life, but the practical life of the best possible kind, τὸ πάντων ἀκρότατον τῶν πρακτῶν ἀγαθῶν (NE 1095a16). The divine ἔργον (it is now evident) is simply the contemplative life, in accordance with NE 1154b26 (διὸ ὁ θεὸς ἀεὶ μίαν καὶ ἀπλῆν χαίρει ἡδονήν) and Mph 1072b14 ff.. The gods are (it appears) the next and possibly only γένος above humanity in the hierarchy of nature. So the contemplative life, while it is not the human ἔργον, is the human *meta-ἔργον*: as we might say, τὸ πάντων ἀκρότατον τῶν ἀγαθῶν ἀπλῶς.

Likewise, we might suggest, the point of Aristotle's remark

that tame animals are superior to wild ones (Po1 1254b10-13) is exactly that tame animals, by their participation in human life, come closer than wild ones to transcending their own γένος. Aristotle's view is, of course, that animals never *actually* achieve this transcendence. But there is a case where something very like this kind of transcendence does, in Aristotle's view, occur: namely, in a child's development into adulthood (NE 1103b14-25).

My contrast between τέλος and meta-τέλος has one further upshot, which brings us back to a more direct attention to the subject of voluntary action. On this view, Aristotle's theory entails that the γένος of humanity (at least) is capable of *making choices* about whether to fulfil its ἔργον, its meta-ἔργον, or neither. In this case, Aristotle is already assuming that humanity is a kind of creature which exhibits voluntary choice. We can now add to this the point that, for Aristotle, the human ἔργον *itself* is voluntary action. For one of Aristotle's several ways of characterising the human ἔργον is in terms of the characteristic human *process* or *movement* (κίνησις). Aristotle says this is uncompelled, contingent, voluntary movement which is up to the agent. Examining this claim, then, will at last enable us to apply what has been established about the human ἔργον to answering the question with which I began this chapter, the question of how to characterise the kind of κίνησις of which humans are typically the ἀρχή.

At NE 1110a15-18, where he is citing evidence for the view that what is done under duress is done voluntarily, Aristotle makes these rather cryptic remarks:

πράττει δὲ ἑκὼν· καὶ γὰρ ἡ ἀρχὴ τοῦ κινεῖν τὰ ὀργανικὰ μέρη ἐν ταῖς τοιαύταις πράξεσιν ἐν αὐτῷ ἐστίν, ὧν δ' ἐν αὐτῷ ἡ ἀρχὴ, ἐπ' αὐτῷ καὶ τὸ πράττειν καὶ μή. (NE 1110a15-18)

Does Aristotle tell us anything new in these words? Certainly the old 'ἀρχὴ inside/ ἀρχὴ outside' contrast is being applied here as a criterion of the voluntary in a way which is familiar from the last chapter. What is new is (i) the gloss on this procedure which follows: ὧν δ' ἐν αὐτῷ ἡ ἀρχὴ, ἐπ' αὐτῷ καὶ τὸ πράττειν καὶ μή; and (ii) the more extended formula ἡ ἀρχὴ τοῦ κινεῖν τὰ ὀργανικὰ μέρη ἐν ταῖς τοιαύταις πράξεσιν ἐν αὐτῷ ἐστίν, in place of the NE's usual ἡ ἀρχὴ ἐν αὐτῷ ἐστίν.

Both these amplifications help clarify Aristotle's meaning. (i) links the enigmatic phrases ἀρχὴ ἐν αὐτῷ and ἐπ' αὐτῷ, 'by me' and 'up to me'. It tells us that Aristotle believes that it is a condition of voluntary action that I should be *able to do otherwise*: that when I do x, I must have been able to do not-x if my doing x is to count as voluntary action. And (ii) focuses our attention on Aristotle's use of two terms which are to be vital in his further elaboration of this claim: namely, the terms ἀρχὴ and κίνησις.

To take these in turn. (i) Ability to do otherwise as a condition of voluntary action appears very clearly in both the NE and the EE. Aristotle would not agree with those moderns who contrast cases of actual compulsion with cases where my action is (they say) causally determined, yet not, in any normal sense, compelled. In such a case, it is often said, I am not *forced* to act as I do- 'it is me acting'- and yet I could not do otherwise. Aristotle's contrast between the allurements of pleasure and the coercion of force (NE 1110b9 ff.) is not meant to be aligned with this modern contrast. Aristotle's contrast makes the exactly opposite point: that pleasures are *not*, normally, causally determining factors, as successful applications of force are; and that only a scoundrel would attempt to equate the two on each and every occasion. The passages of the EE (discussed in Ch.1) in which Aristotle argues that pleasures can *sometimes* compel, give further evidence that he believed that pleasures do not *normally* compel.

Aristotle's express view is that, if *for any reason* I could not have done otherwise than I do, I do not act voluntarily. In Aristotle's view, if determinism is true, then *all* action is compelled. So, at NE 1113b7-10, Aristotle makes ability to do otherwise a condition of moral responsibility *because* it is a condition of voluntary action:

ἐφ' ἡμῖν δὴ καὶ ἡ ἀρετὴ. ὁμοίως δὲ καὶ ἡ κακία· ἐν οἷς γὰρ ἐφ' ἡμῖν τὸ πράττειν, καὶ τὸ μὴ πράττειν, καὶ ἐν οἷς τὸ μὴ, καὶ τὸ ναί· ὥστ' εἰ τὸ πράττειν καλὸν ὄν ἐφ' ἡμῖν ἐστί, καὶ τὸ μὴ

πράττειν ἐφ' ἡμῖν ἔσται ἀλοχρὸν ὄν... (NE 1113b7-10)

Compare EE 1225b8: ὅσα μὲν οὖν ἐφ' ἑαυτῷ ὄν μὴ πράττειν πράττει μὴ ἀγνοῶν καὶ δι' αὐτόν, ἐκούσια ταῦτ' ἀνάγκη εἶναι...

(ii) This view, that voluntary action is necessarily contingent, is spelled out by what Aristotle has to tell us about the meanings of the terms ἀρχὴ and κίνησις. Here the exiguous argument of NE 1110a15-18 can be filled out by reference to EE 1222b15-1223a20 (cp. MM 1187a30-b30).

EE 1222b16 tells us that εἰσὶ δὲ πᾶσαι μὲν αἱ οὐσίαι κατὰ φύσιν τινὲς ἀρχαί: 'all substances are, by nature, principles of some sort'. To be an ἀρχὴ of some type *x*, apparently, involves the ability to produce other ἀρχαί of that type *x* (b17-18). But there is another sense of ἀρχὴ which is more important (b21-22: τῶν δ' ἀρχῶν ὅσαι τοιαῦται, ὅθεν πρῶτον αἱ κινήσεις, κύριαί λέγονται). In this sense of ἀρχὴ, what is humanity the ἀρχὴ of?

ὁ δ' ἄνθρωπος ἀρχὴ κίνησεώς τινος· ἢ γὰρ πράξις κίνησις. (EE 1222b28-30)

ὁ γ' ἄνθρωπος καὶ πράξεων τινῶν ἐστὶν ἀρχὴ, μόνον τῶν ζῴων· τῶν γὰρ ἄλλων οὐθέν ἐπινοοῦμεν ἄν πράττειν. (EE 1222b19-21)

(In spite of this latter remark, Aristotle himself sanctions the use of πράττειν for animal behaviour at NE 1111a26-27: what is done through θυμὸς or ἐπιθυμία should not be called

'involuntary', because the undesirable consequence would be that οὐδέν ἔτι τῶν ἄλλων ζῶων ἐκούσιως πράξει. We have three names for two kinds of action, apparently: τὸ ἐκούσιον, πράξις, action on προαίρεσις. Aristotle is quite clear that τὸ ἐκούσιον does not equal action on προαίρεσις (NE 1111b7-8), but less clear about whether πράξις equals τὸ ἐκούσιον or action on προαίρεσις.)

What, then, does it mean for humanity to be the ἀρχὴ of πράξεων τινῶν, which are constituted by κινήσεις? And, even if we can't properly speak of animal πράξις, is there anything in their case corresponding to this kind of being an ἀρχή?

An obvious answer to the first question, given the last two quotations, is also suggested by NE 1110a15, as above: ἡ ἀρχὴ τοῦ κινεῖν τὰ ὀργανικὰ μέρη ἐν ταῖς τοιαύταις πράξεσιν ἐν αὐτῷ ἐστίν. For a human to be the ἀρχὴ of πράξεων τινῶν is, at least, for her to be the ἀρχὴ τοῦ κινεῖν τὰ ὀργανικὰ μέρη. The human is ἀρχὴ of some particular πράξις if and only if the κινήσεις, the movement(s) of, or change(s) in, her body which that πράξις involves, are causally originated by herself. This answer is supported when Aristotle continues the EE passage with (1222b21): τῶν δ' ἀρχῶν ὅσαι τοιαῦται, ὅθεν πρῶτον αἱ κινήσεις...

This answer in turn suggests an answer to the second question, about animals. There is a correspondent to human πράξις in animals, even if their behaviour cannot be called πράξις. To put it noncommittally, an animal is, in the proper sense, an ἀρχὴ of some particular κίνησις if and only if that κίνησις is

causally originated by itself. This form of the statement, indeed, could apply to human beings too. For- and this is the point- *πρᾶξις* is here being treated by Aristotle as *necessarily involving* κίνησις (in stark opposition, of course, to those places where Aristotle insists that *πρᾶξις*, as opposed to *ποίησις*, is an ἐνέργεια, and not a κίνησις: on which v. Ackrill (1965).

Having set the animals and their typical κινήσεις *below* humanity, Aristotle next sets the divine and its typical κινήσεις *above* humanity:

τῶν δ' ἀρχῶν ὅσαι τοιαῦται, ὅθεν πρῶτον αἱ κινήσεις, κύριαί λέγονται, μάλιστα δὲ δικαίως ἀφ' ὧν μὴ ἐνδέχεται ἄλλως, ἦν ἴσως ὁ θεὸς ἄρχει. (EE 1222b21-23)

The divine characteristically originates *necessary*, non-contingent κινήσεις; Aristotle has in mind both θεωρία and the perfect circular motion of the 'fixed stars' (Mph 1072a7-b31). Animals characteristically originate κινήσεις which are unworthy of the name of *πρᾶξις*, or at any rate of the name of *προαίρεσις*; elsewhere Aristotle tells us that animals κινεῖται ἀορίστως (dA 434a4), move without defined (self conscious) purpose. What kind of κίνησις will humans characteristically originate when they originate *πρᾶξις*, given that (as we have seen) Aristotle takes seriously the thought that humanity is set *between* animals and gods in the hierarchy of nature?

Aristotle argues that *πρᾶξις* is a kind of *κίνησις* which, unlike the animals' *κίνησις*, is not (in some sense or other) sub-purposive; and which, unlike the divine *κίνησις*, *ἐνδέχεται γενέσθαι ἐπὶ τᾶναντία* (1223a2). Aristotle's argument for this conclusion is entertainingly opaque. It seems to turn on a geometrical analogy. We have what we call *ἀρχαὶ* in geometry, although this name is only given analogously, and is not strictly appropriate to such principles (b23-25). If these '*ἀρχαὶ*' are changed, so, necessarily, is every consequence that follows from them (b25-27, b33-37). This is the point of analogy with genuine *ἀρχαὶ* which justifies the geometrical principles' being called *ἀρχαὶ* too. The points of *disanalogy* with genuine *ἀρχαὶ* are (i) that mathematical *ἀρχαὶ* do not *move* anything (b23-25), and (ii) that mathematical *ἀρχαὶ* do not 'admit of coming to be otherwise' (1223a1).

What we may learn from the geometrical analogy is this: that *ἀρχαὶ* of any given kind (as it might be, contingent or non-contingent) necessarily give rise to *κινήσεις* of the same kind (1222b41-1223a3, following Kenny (1979)'s interpretation of *ἐντεῦθεν* at 1223a2). By contrast with the mathematical *ἀρχαὶ*, it follows that the genuine *ἀρχαὶ* (i) *do* produce actual *κινήσεις*, and (ii) produce *κινήσεις* which *do* 'admit of coming to be otherwise'. Aristotle expects us to find it obvious that humans are contingent *ἀρχαὶ*, from which it follows (i) that humans are genuine *ἀρχαὶ*, and (ii) that humans characteristically produce *κινήσεις* which *do* 'admit of coming to be otherwise'. Since Aristotle has also said (EE 1222b29-30) that *πρᾶξις* is the

κίνησις characteristic of humans, it follows, further, that πρᾶξεις are κινήσεις which 'admit of coming to be otherwise'.

Human πρᾶξις, then, is a κίνησις which can be placed in the middle of a range of κινήσεις of different degrees of contingency, between the products of chance and the animals' κίνησις on the one hand, and the divine κίνησις on the other. The idea of a hierarchy of nature is obviously at work here. Compare Aristotle's remarks about the domain of deliberation, NE 1112a18-b11. As Joachim puts it (Joachim 1955, p.101):

'Aristotle excludes not only τὰ ἀίδια (the timeless and unchangeable) and τὰ ἀεὶ κατὰ ταῦτα γινόμενα (absolutely regular changes), but also the haphazard (τὰ ἀπὸ τύχης)- changes which follow no rule at all, whether outside (τὰ ἄλλοτε ἄλλως) or within the region of human purposive action. He thus restricts deliberation to τὰ ὡς ἐπὶ τὸ πολὺ (things that happen in a certain way for the most part)'. And it is in our dealings with τὰ ὡς ἐπὶ τὸ πολὺ that we are most characteristically human, and fulfil our ἔργον.

To repeat a point I made at the outset: Aristotle argues *conditionally*, not categorically, against determinism. Thus such passages as I have considered here will suggest this argument against the non-existence of human moral responsibility:

1. A does X freely only if A does X voluntarily.

2. For any agent P and any voluntary action, Q, done by P, Q is not a voluntary action unless P could have done otherwise than Q.
3. Therefore, if X was a free action, it cannot have been the case that A could not have done otherwise than X.

Aristotle often fills out his claim that it is an unappealing position to say that there is no freedom of action, or no moral responsibility, by pointing out the dependence of the notion of free (or morally responsible) action on that of voluntary action. In his remarks about κίνησις he goes a step further, by pointing out the dependence of the notion of voluntary action on that of contingency, another notion which, it would seem to him, it would be very outlandish to deny. And his notion of contingency dovetails with something else again, namely his notion of the hierarchy of nature, into which human beings fit precisely by being such creatures as to produce contingent movements. Aristotle's tactics suggest that he thinks that anyone who denies any one of these notions, must deny them all. Perhaps we may say that Aristotle's arguments about κίνησις are dialectical (in the strict sense: see §2) because he thinks that the relations of contingency, voluntariness, freedom and responsibility are a matter of definitions. For some action of mine to be voluntary is for it to be 'up to me to do or not to do'; for it to be 'up to me to do or not to do' is for it to be a κίνησις which 'admits of coming to be otherwise'; for it to be a κίνησις which 'admits of coming to be otherwise' is for there to be real contingency in the world. This dialectical argument that there really is such contingency is something which Aristotle pursues with conviction

and spirit- not just in the ethical writings (cp. dI 9).

For an action to be uncompelled, then, is for it to be a κίνησις satisfying the details of the above description. It is also, to conclude this section, for that action to involve a κίνησις which fits under the third or fourth category in Table One (p.85), which sets out, as fully as may be, the array of the hierarchy of nature and the different κινήσεις which are, sometimes, called the ἔργα of its different γένη. (N.B. that, in this table, as above, 'locus in nature' means, not sole or chief locus in nature of the κίνησις in question, but the point in the hierarchy of nature at which it is first characteristically present. Note also that 2c subsets 2b, and 2b subsets 2a.)

2. Aristotle's Epistemology

The negative requirement that a voluntary action should be uncompelled has for Aristotle the corollary that it should be of the kind described above. What description(s), in turn, must be satisfied as a corollary of the negative requirement that a voluntary action should not be done in ignorance?

Aristotle may seem to have less to say about this question than about the corresponding question which I have just dealt with, regarding compulsion. In fact, one prominent contemporary Aristotelian has actually suggested that, from an Aristotelian

Table One: The Hierarchy of Nature

ΚΙΝΗΣΙΣ	MENTAL STATE NEEDED	LOCUS
	(a: for performance of this κίνησις; b: for understanding it)	(a: in Ar.; b: in nature)
1. κίνησις [μὴ ἐνδεχόμενη τοῦ ἄλλως ἔχειν]	a: θεωρία? b: νοῦς/ἐπιστήμη	a: dA 404b6-8, dC 1.2-4 b: the stars
2a. κίνησις ἐνδεχόμενη τοῦ ἄλλως ἔχειν	a: ? b: νοῦς/ἐπιστήμη	a: dA 404b6-8, EE 1224b41 ff. b: sublunary world
2b. τὸ ἑκούσιον/πράξεις	a: ὄρεξις + φαντασία b: ?	a: dA 433b32ff. and EE 1223a21 b: animals, children
2c. πράξεις/προαίρεσις	a: βουλευσίς/φρονήσις b: ?	a: NE 3.3, 6.9, EE 2.7-10 b: adult men
[3. θεωρία	a: σοφία b: ?	a: NE 6.7, 10.8, 1154b26; Mph 1072b14 ff. b: excellent men and gods]

point of view, 'the whole epistemological turn of philosophy is the outcome of a mistake' (MacIntyre 1990, p.69). It is, indeed, widely supposed that the commonsensical Aristotle would have had no more patience than MacIntyre evidently has with epistemological questions, at least those in the form made classical by Descartes (and Pyrrho). Undoubtedly, Aristotle has little to say about such questions explicitly. However, this does not mean either that these questions do not interest him, or that he has no way of dealing with them.

On what basis could it be supposed that Aristotle had no patience with, or aptitude for, epistemological questions? Perhaps this claim would come as part of the wider thesis that Aristotle is a foundationalist, and so a commonsense realist in epistemology. But that wider thesis is simply false: as I will argue, Aristotle is *not* a foundationalist. Or perhaps it could be argued that Aristotle's procedure is not to overthrow common sense realism, but to buttress it. However, Aristotle's attitude to what he takes to be the 'appearances' or 'common opinions' is not automatically uncritical; anything but. Or, thirdly, this anti-epistemological reading of Aristotle might be argued for on the basis of a passage like Mph 1008b13-18. But, in the first place, this passage concerns a logical principle, not an epistemological one; and in the second place, the dismissive remarks of Mph 1008b13-18 come after three pages of very careful statement of and argument for the principle in question (1005b18 ff.), in the course of which, moreover, Aristotle actually suggests ways of *arriving at* that principle. There are, then, no

good general reasons (and no specific ones, either) for supposing Aristotle to be uninterested by epistemological questions.

But, if Aristotle has an epistemology, what is it? That question is rather difficult to answer head on; because the point in Aristotle's philosophy at which his epistemological method becomes clearest is not in his response to *epistemological* scepticism, but in his response to *moral* scepticism. Accordingly, I shall deal with Aristotle's general epistemology only *via* his moral epistemology. Both kinds of epistemology are, of course, relevant to the topic of voluntary action and responsibility in Aristotle. Broadly, moral epistemology gives us an account of the possibility of knowledge of principles, lack of which does not hinder voluntary action and is blameworthy; general epistemology gives us an account of the possibility of knowledge of particulars, lack of which hinders voluntary action and, normally, exculpates.

At first sight, Aristotle's tentative statements in definition of virtue, justice and related terms seem— as he himself admits— to succumb to a certain endemic *redundancy*:-

Ἄλλ' ἕως τὴν μὲν εὐδαιμονίαν τὸ ἄριστον λέγειν ὁμολογούμενόν τι φαίνεται... (NE 1097b22)

ὑπόκειται ἄρα ἡ ἀρετὴ εἶναι ἡ τοιαύτη περὶ ἡδονᾶς καὶ λύπας τῶν βελτίστων πρακτικῆ, ἡ δὲ κακία τοῦναντίον. (NE 1104b27-28)

ὁ μὲν οὖν ἄ δεῖ καὶ οὐ ἔνεκα ὑπομένων καὶ φοβούμενος, καὶ ὡς
δεῖ καὶ ὅτε, ὁμοίως δὲ καὶ θαρρῶν, ἀνδρεῖος. (NE 1115b18-19)

The good is hard to define (cp. NE 1094b14 ff.). But this is no surprise, since, as Aristotle is careful to say several times, virtue itself is *difficult*:

περὶ δὲ τὸ χαλεπώτερον αἰεὶ καὶ τέχνη γίνεται καὶ ἀρετή· καὶ γὰρ τὸ εὖ βέλτιον ἐν τούτῳ. (NE 1105a8-10)

διὸ καὶ ἔργον ἐστὶ σπουδαῖον εἶναι· ἐν ἐκάστῳ γὰρ τὸ μέσον λαβεῖν ἔργον. (NE 1109a25)

οἱ δ' ἄνθρωποι ἐφ' ἑαυτοῖς οἴονται εἶναι τὸ ἀδικεῖν, διὸ καὶ τὸ δίκαιον εἶναι ῥάδιον, τὸ δ' οὐκ ἔστιν... (NE 1137a5-6)

Aristotle famously remarks that 'the same kind of accuracy is not to be sought in all arguments alike; in any particular kind of matter, an educated person looks for only that degree of precision which its nature permits' (NE 1094b13). But 'practical matters and questions of interests contain no fixed matter' (NE 1104a4). So is it his policy to refuse to offer us any definite moral epistemology at all? Apparently not. There are several notions, prominent (or supposedly prominent) in his thought, which might seem obvious candidates for a key role in his moral epistemology. As a closer look at them will now show, none of these 'obvious' candidates is in fact adequate to the job. Yet it is their very inadequacies which will lead us to a possibly less

obvious, but certainly more suitable, candidate for the role of criterion of the good.

Six candidates for the role of criterion of the good are given by these suggestions about how Aristotle thinks we should do moral epistemology.

a. *Functional naturalism*: the basis of ethics is given by reference to the concept of 'function' already discussed. (NE 1097b21-1098a18)

b. '*Aurea Mediocritas*': the basis of ethics is given by defining a *mean* between two extremes. (NE 2.6-9, 3.6-5.4)

c. *Dogmatism*: the basis of ethics is given in a set of specific, fixed moral principles or rules.

d. *Principle foundationalism*: the basis of ethics is given in those universal moral principles (ὀρθος λόγος, NE 1103b34, 1144b27) the 'immediate perception' or 'intuition' of which (θεωρεῖν, NE 1140b10) furnishes an end to moral reasoning.

e. *Hedonism*: the basis of ethics is given by asking 'What is pleasant?'; or at least the presence of the pleasant is a necessary condition of the presence of the good. (NE II.3)

f. *Particular foundationalism*: the basis of ethics is given in those particular moral truths the 'immediate perception' or 'intuition' of which (θεωρεῖν, NE 1140b10) furnishes an end to moral reasoning. (So, supposedly, APo I.1 and NE 1139b25-36, 1140b13-20, 1143a25-b6.)

My arguments for rejecting each of these alternatives will

lead me to make my own suggestion about how Aristotle thinks we should do moral epistemology; a suggestion which I will then apply to Aristotle's treatment of general epistemology.

2a. Functional Naturalism

The concept of function has indeed— as I have argued in §1— an important role in Aristotle's moral philosophy. But it does *not*, properly speaking, play the role of *definiens* of what is good. For it leads us to this dilemma: should we interpret the concept loosely, using it as no more than a rule of thumb? Or in a rigorous, specific way, so that we take it that by using this concept we can arrive at moral verdicts on particular cases which can be deduced from the nature of those cases? Any loose interpretation is liable to be unhelpful in specific cases, and so not much use. But any rigorous interpretation has to justify itself in face of the fact that there are plenty of *other* conceivable ways of tightening the doctrine up enough for it to give verdicts on specifics. What gives *one* formalisation any kind of priority over any other?

Moreover, Aristotle makes just as much of several other approximating methods of delineating what is good as he does of the concept of function. The concept of function is not clearly prior to any of these. In particular, it is not obviously prior to the notion of the mean. So what grounds are there for saying that the doctrine of the mean is to be understood in the terms of

the concept of function, rather than *vice versa*?

2b. 'Aurea Mediocritas'

The same points, *mutatis mutandis*, apply to the doctrine of the mean. This doctrine, important to Aristotle though it is, cannot provide the *definiens* of what is good, because of the loose, general and inconclusive nature of any *pretheoretical* conception of the mean, and the controversial nature of any *posttheoretical* conception of the mean. Also, the doctrine of the mean is no more clearly prior to the concept of function than *vice versa*.

The main deficiency which (2a) and (2b) share, as candidates for the role of criterion of the good, is suggestive. It is that both notions are of such loose content, so hard to put to precise use, that to make real sense of them we need the further notion of an expert in using them. More of which below.

2c. Dogmatism

It cannot be true for Aristotle that the basis of ethics is given in a set of specific, fixed moral principles or rules. This is for the simple reason that, as he explicitly says, Aristotle does not believe there *are* any such fixed moral principles or rules.

If he *is* a dogmatist in ethics, it should not be hard to find a case of a moral rule which he says is indefeasible. But there is only one passage which might be taken this way in the whole of his ethical writings, NE 1110a26-29:

ἔνια δ' ἴσως οὐκ ἔστιν ἀναγκασθῆναι, ἀλλὰ μᾶλλον ἀποθανατέον παθόντι τὰ δεινότατα.

Here the indefeasible moral rule would, presumably, be 'No matricide' (a28-29). But note, first, that 1110a26 ff. is not an unqualified claim: ἴσως is Aristotle's usual word for expressing the tentative, provisional nature of a remark. Second, note the context of this remark, in a discussion, *not* of what is always right and what is always wrong, but of ποῖον ἀντὶ ποίου αἰρετέον, καὶ τί ἀντὶ τίνος ὑπομενετέον (1110a30), that is, of *which actions are to be preferred to which* (and/ or, following Joachim (1951, p.98), what are the pros and cons involved in any particular action).

Aristotle is not making the dogmatist type of claim which, say, Geach (Geach 1977) would want to make- in this case, the claim that matricide is a categorically wrong action. His point is that matricide is not, for a good person, a practically conceivable *alternative* even to such extreme actions as allowing oneself to be tortured to death. If one is to be a virtuous person, and not an unprincipled coward or hypocrite, there must (Aristotle believes) be some things for which one is prepared to die. This does not mean that anything is absolutely forbidden in

the sense which the dogmatist needs. Notice that, even when talking of an act for which there could, in his view, never be any easily imaginable justification, Aristotle still uses the language of *alternatives* and not that of *rules*. The idea that ethics could be reduced to rules is a foreign one to him.

In any case, there is good evidence that no moral rules whatever are, for Aristotle, actually indefeasible. In NE V, Aristotle says that there is a difference between τὸ δίκαιον, *justice proper* (as expressed in the character and actions of just persons, NE 1129a7-10), and τὸ δίκαιον κατὰ νόμον, *equity* (state justice, as expressed in the laws of a state: NE 1137b12). The difference is that equity, unlike justice proper, is hindered in its attempts to make settlements expressive of τὸ ἐπιεικὲς ('what is fair') by the fact that equity is expressed in laws, which are, by their very nature, *exceptionless rules*:

ὁ μὲν νόμος καθόλου πᾶς, περὶ ἐνίων δ' οὐχ οἷόν τε ὀρθῶς εἰπεῖν καθόλου. (NE 1137b13-15)

Exceptionless rules- Aristotle explicitly says here- are bound to be defective. The reason is (NE 1137b20) the variable nature of their subject matter: παραλείπει ὁ νομοθέτης καὶ ἤμαρτεν ἀπλῶς εἰπὼν (b22). (Compare NE 1107a29-32: ἐν γὰρ τοῖς περὶ τὰς πράξεις λόγοις οἱ μὲν καθόλου κοινότεροί εἰσιν, οἱ δ' ἐπὶ μέρους ἀληθινώτεροι; and NE 1094b19-22, 1142a18-19.) Hence we do not reach τὸ ἐπιεικὲς by simply accepting the universal rules of τὸ δίκαιον κατὰ νόμον but, on the contrary, by *looking for the*

grounds, if any, on which an exception should be made to that rule:

καὶ ἔστιν αὕτη ἡ φύσις ἢ τοῦ ἐπιεικοῦς, ἐπανόρθωμα νόμου ᾧ ἔλλείπει διὰ τοῦ καθόλου. (NE 1137b27)

From all this it seems clear that Aristotle's moral theory is not based upon indefeasible moral rules. On the contrary, any such rule could only be of use to us in moral discernment when we know already whether the case before us is one in which we should simply apply the rule, or else make an exception to it, and if so of what kind. If, in ethical theory, we wish to invoke even the notion of *defeasible* moral rules- putting aside the implausible idea of *indefeasible* ones- we will need the further notion of an expert in using or applying these rules.

Notice that, when Aristotle's argument leads him to invoke the idea of defeasible rules in equity, this further notion of the expert comes in at once. For (he tells us at NE 1137b20-24) the right response, when one becomes aware of an exception to the legal rule, is to rectify the anomaly by doing '*what the lawgiver himself would have done had he been present, and would have enacted if he had known*'. The expert in equity, then, is the νομοθέτης.

2d. Principle foundationalism

Nor can it be true for Aristotle that the basis of ethics is given in universal moral principles, the 'immediate perception' or 'intuition' of which furnishes an end to moral reasoning. This in spite of what a naive reading of, for example, NE 1140b17-20 might suggest:

αἱ μὲν γὰρ ἀρχαὶ τῶν πρακτῶν τὸ οὐ ἕνεκα τὰ πρακτά· τῷ δὲ διεφθαρμένῳ δι' ἡδονὴν ἢ λύπην εὐθύς οὐ φαίνεται ἀρχή, οὐδὲ δεῖν τοῦτου ἕνεκεν οὐδὲ διὰ τοῦθ' αἰρεῖσθαι πάντα καὶ πράττειν. (NE 1140b17-20)

This, however, is not because Aristotle does not believe there are any moral first principles. It is because, although he admits the existence of something like such principles, those principles do not play the role of providing a metaphysical long-stop for his ethics; nor are they the object of 'immediate perception' or 'intuition'.

Aristotle, I have argued, is not a dogmatist, for there are, in his view, no actually indefeasible moral rules. If this is right, then clearly we cannot accept the claim that Aristotle's 'moral principles' are anything like such rules. His moral principles are not such that, once we have worked back to them, no further working-back is possible. What then are his 'moral principles'? And what role do they serve in his moral theory? This is a big question, and one to which Terence Irwin has

devoted what I think is his *magnum opus* (Irwin 1987). Here I can only venture brief remarks.

Aristotelian moral principles are in important ways analogous to other kinds of argumentative principle. But Aristotle's general argumentative principles are not what a principle foundationalist would want them to be. Despite the apparent sense of APo 72a6-7, b18-24, Aristotle's principles are not utterly indisputable and prime starting points for *all* argument. Rather, they are points at which *one kind* of argumentative method is exchanged for another, namely 'induction' for 'deduction':

μη λανθανέτω δ' ἡμᾶς ὅτι διαφέρουσιν οἱ ἀπὸ τῶν ἀρχῶν λόγοι καὶ οἱ ἐπὶ τὰς ἀρχάς... (NE 1094a31-32)

ἡ μὲν γὰρ ἐπαγωγή ἀρχῆς ἐστὶ καὶ τοῦ καθόλου, ὁ δὲ συλλογισμὸς ἐκ τῶν καθόλου. εἰσὶν ἄρα ἀρχαὶ ἐξ ὧν ὁ συλλογισμὸς, ὧν οὐκ ἔστι συλλογισμὸς· ἐπαγωγή ἄρα. (NE 1139b28-31)

Thus Aristotelian first principles, so far from being objects of 'immediate perception' or 'intuition' as they are in intuitionism or foundationalism, are themselves *results of a prior logical process*. There can, paradoxical as it may sound, be argument *to* first principles (ἐπαγωγή), as well as argument *from* those first principles (συλλογισμὸς or λόγος). Note that both συλλογισμὸς and λόγος, besides meaning 'deduction' as opposed to 'induction', can also mean 'argument' in general: so NE 1094a32 above, and APr 68b30. This, no doubt, is one root of the mistaken

view that Aristotelian 'induction' is not a *pukka* form of argument. It also helps explain the notorious problem about *προαίρεσις*, that sometimes (e.g. 1112b13) it has to do only with means, and at other times seems to be of ends too (e.g. 1139a33). Practical reasoning too may consist both of inductive and deductive parts. Strictly, *προαίρεσις* is practical reasoning from first principles, but the word is sometimes more loosely used by Aristotle to suggest practical reasoning *to* first principles: v., e.g., NE 1144a20.

In what, then, does argument *to* first principles consist? One of Aristotle's answers is that *ἐπαγωγή* is equivalent to *definition*, *ὀρισμός*: 'Definition is a "thesis" or a "laying something down"', according to APo 72a21. APo 72b23-24 tells us that, besides scientific knowledge, there is its originative source, which enables us to recognise the definitions. There is also NE 1142a26: *ὁ μὲν γὰρ νοῦς τῶν ὄρων, ὃν οὐκ ἔστι λόγος*. Indeed, Aristotle has a whole science of definition, laid out in beautiful detail in the *Topics*, Bks. VI and VII, which I suggest gives further detail on what kinds of argumentative processes can be prior to and result in the positing of any first principle.

Argument *to moral* (or, better, practical) first principles will be of an analogous kind to any other argument to first principles. It will start from what is uncertain to the reasoner and proceed to what is more certain, just as deduction proceeds in the opposite direction. An example would be when a number of non-cogent considerations, taken together, yield a more cogent

overall view; this overall view then becomes the 'first principle' of deduction. Thus: 'I have a reason to do x, and a reason to do y; but my reason to do y outweighs my reason to do x; so I have more reason to do y' is inductive reasoning; and a deductive reasoning, about how to achieve y, may then follow on from 'I have more reason to do y', taking this as its first principle. For more on this, v. Ch.3.

To say this much is already to make it plain that the 'first principles' of moral discourse are in no foundationalist sense the *foundations* or *basis* of that discourse. The role of the 'first principles' of moral discourse is the same as that of the first principles of any other kind of discourse: to provide a basis for συλλογισμός in that discourse (i.e., in this case, for *practical* reasoning). It is not to be the epistemological warrant of the whole discourse; that job is performed elsewhere.

However, saying that the first principles of moral discourse are *definitions* prompts the question 'Definitions of what?'. My answer turns on a point noticed in my exposition of the ἔργον argument. To define a τέχνη is itself, I said, to define the τέλος of that τέχνη: 'To define what performing a function (ἔργον) is, just is to define what performing that function *well* is'.

So the ἀρχαὶ of ethics may be found within the correct understanding of those τεχναὶ and other human practices excellence in which, on Aristotle's view, is constitutive of

human excellence in general. To have a correct understanding of what is required of me, in the context of any particular pursuit in which I engage, is no different from having a correct definition of what that pursuit itself *is*. (This is one point made by Aristotle's remark αἱ μὲν γὰρ ἀρχαὶ τῶν πρακτῶν τὸ οὐ ἕνεκα τὰ πρακτά (NE 1140b17).) Thus my definition of what pursuit it is that I am engaged in is itself the first principle of my action within that pursuit.

Note the consequence: if, in ethical theory, we wish to invoke the notion of the 'moral first principle', given by the definition of the τέχνη in which one is involved, we will need the further notion of a prior *expertise in definition*. Once again, we come back to the notion of the expert.

2e. φρόνησις and Particular Foundationalism

Again, it is not true that the basis of ethics for Aristotle is given in those particular moral 'perceptions' which furnish another kind of end to moral reasoning. This, again, is not because he does not believe there are any such 'perceptions'. It is because, although he admits the existence of something like such 'perceptions', they do not play the role of providing an ultimate basis to his ethics.

My phrase 'particular moral perceptions' is suggested by what Aristotle has to tell us about φρόνησις ('practical

intelligence'). φρόνησις, he says, is of 'the ultimate particular, for what is to be done is an ultimate particular' (τοῦ γὰρ ἐσχάτου ἐστὶν [φρόνησις]· τὸ γὰρ πρακτὸν τοιοῦτον, NE 1142a25-26). He then draws an analogy between φρόνησις and νοῦς (normally translated 'scientific intelligence'). According to NE 1141b3, νοῦς is one of the two components- the other is ἐπιστήμη ('scientific knowledge')- of Aristotelian σοφία, the subject matter of which seems to be a heady mix of theology and astronomy. The point of the analogy, then, is that φρόνησις is more like νοῦς than it is like ἐπιστήμη (a24-25).

¶ 1141b3

So what is the contrast between νοῦς and ἐπιστήμη? The answer is, apparently, that νοῦς is the ability to recognise success in working, inductively, *towards* the first principles of science; ἐπιστήμη is the result of working, deductively, *from* those first principles. νοῦς is the ability to see that a piece of ἐπαγωγή or ὀρισμός has or has not been a success, whereas ἐπιστήμη consists in the making of συλλογισμοὶ (so NE VI.6). But this is not the only sense in which Aristotle uses νοῦς:

both NE

¶ 1143a33-35 καὶ ὁ νοῦς τῶν ἐσχάτων ἐπ' ἀμφοτέρα· καὶ γὰρ τῶν πρώτων ὄρων καὶ τῶν ἐσχάτων νοῦς ἐστὶ καὶ οὐ λόγος. (NE 1143a33-35)

¶ 1143a33-35

¶ 1143a33-35 Kenny writes: 'In a theoretical discipline like geometry, according to Aristotle, [νοῦς] comes in at two points: first, in the appreciation of first indemonstrable principles or axioms; secondly, in the realisation that particular individuals are instances of generalisations' (Kenny 1979, p.151). The ultimate

particulars, τὰ ἐσχάτα, of which there is also νοῦς, καὶ οὐ λόγος, are *particular facts*, concerning which νοῦς is just seeing that 'so and so is the case' (NE 1095b7). νοῦς in this sense is a kind of αἴσθησις: τοῦτων [sc. τῶν καθ' ἕκαστα] ἔχειν δεῖ αἴσθησις, αὕτη δ' ἐστὶ νοῦς (NE 1143b5-6).

(To make things more confusing, Aristotle has a double sense of ἀρχή as well as a double sense of νοῦς. Sometimes, as we have seen (§2d), he calls the understanding of right ends the ἀρχή: so 1140b17 αἱ μὲν γὰρ ἀρχαὶ τῶν πρακτῶν τὸ οὐ ἕνεκα τὰ πρακτά. But at other times he describes the αἴσθησις of ultimate particulars as the ἀρχή of practical reasoning: so ἀρχή γὰρ τὸ ὅτι (1095b7), and 1143b4 ἀρχαὶ γὰρ τοῦ οὐ ἕνεκα αὗται. The answer to this puzzle is, no doubt, that speaking more accurately neither the understanding of right ends nor the αἴσθησις of ultimate particulars is *the* ἀρχή of practical reasoning; *each*, in a different way, is *an* ἀρχή of practical reasoning. Both are necessary for practical reasoning; only together are they, possibly, necessary and sufficient. Aristotle makes reference to both when he is emphasising different aspects of how action on practical reasoning arises.)

In any case, it is presumably to the νοῦς of *ultimate particulars* that φρόνησις is analogous, *not*, as NE 1142a26-27 seems to suggest, the νοῦς of *first principles*:

ὁ μὲν γὰρ νοῦς τῶν ὄρων, ὧν οὐκ ἔστι λόγος, ἢ [φρόνησις] δὲ τοῦ ἐσχάτου, οὐ οὐκ ἔστιν ἐπιστήμη ἀλλ' αἴσθησις. (NE 1142a26-27)

NE 1143b2-4 suggests that φρόνησις is the knowledge (νοῦς) of the 'contingent particular fact' (τοῦ ἐσχάτου καὶ ἐνδεχομένου, b3) which gives the practical syllogism its particular premiss. That is: φρόνησις is the ability to see that this particular case is covered by the general rule of the universal premiss; the ability to see which are the morally relevant features of a case and which are not.

But any case in which a moral judgement is called for may have indefinitely many different features which might be thought morally relevant: πολλαὶ γὰρ διαφοραὶ εἰσὶν ἐν τοῖς καθ' ἕκαστα (1110b9). Can the right features, those to which one *ought* to attend, simply be 'read off' the case by any means— say by attributing a privileged status to certain sorts of moral sentiment or 'perceptions' of moral quality? Aristotle clearly thinks not. To provide such a means would be to provide a *general rule* for *particular cases*. To say that 'This sort of "gut feeling" is *always* reliable' would be to return, even if in a covert form, to moral dogmatism. (It would also land us back in 'private language' difficulties about the reidentification of 'internal' experiences.) No kind of moral intuition, however strong or distinctive, is necessarily right (why *should* it be?). The real job of the ethicist, much harder but much more worthwhile, is to tell us how to develop, not 'gut feelings', but discernment about (*inter alia*) one's own 'gut feelings' and one's perceptions of what count as morally relevant features of a case and what aren't. This kind of discernment, this learning to see one's way through the inexhaustible complexities of moral

questions, is an *expertise*; and hence the notion of φρόνησις too is useless without the prior notion of an expert in the application of φρόνησις.

In any case, if Aristotle's claim about φρόνησις is that φρόνησις is the ability to see that *this general principle applies* in any particular case, this means that there is *no* φρόνησις, unless one's general principles are in good order. There is only δεινότης (NE 1144b15 ff.), for φρόνησις is characteristic only of the good person (NE 1144a29-31). ἡ μὲν γὰρ ἀρετὴ τὸν σκοπὸν ποιεῖ ὀρθόν, ἡ δὲ φρόνησις τὰ πρὸς τοῦτον (NE 1144a7): φρόνησις either works in harness with ἠθικὴ ἀρετὴ or not at all (NE 1144b30-32). φρόνησις is only 'the ability to see what is right' in a particular case, if the general principle which it applies to that case *is already describable as good*. Therefore, since the existence of φρόνησις is logically dependent on the goodness of moral principles which (I have argued) themselves cannot be the basis of Aristotle's ethics, *a fortiori* φρόνησις cannot be the basis of Aristotle's ethics either.

2f. Hedonism

The grounds for calling Aristotle an ethical hedonist would be given by remarks like NE 1104b9-10 (περὶ ἡδονὰς γὰρ καὶ λύπας ἐστὶν ἡ ἠθικὴ ἀρετὴ) and, almost identically, EE 1220a35 (περὶ ἡδέα καὶ λυπηρὰ καὶ ἡ ἀρετὴ καὶ ἡ κακία). Such remarks, it might

seem, are backed up by Aristotle by a general hedonism in philosophical psychology:

καὶ τὸ διώκειν δ' ἅπαντα καὶ θηρία καὶ ἀνθρώπους τὴν ἡδονὴν
σημεῖόν τι τοῦ εἶναι πῶς τὸ ἄριστον αὐτήν. (NE 1153b25-31; cp.
1172b35 ff.)

Moreover, Aristotle says quite clearly that pleasure is an essential part of the Good Life:

πάντες τὸν εὐδαίμονα ἡδὺν οἴονται βίον εἶναι, καὶ ἐμπλέκουσι
τὴν ἡδονὴν εἰς τὴν εὐδαιμονίαν, εὐλόγως. (NE 1153b14-16)

τὸ τε ἀπλῶς ἀγαθὸν καὶ ἡδὺ ἀπλῶς ἐστίν. (NE 1156b23)

Does it follow, then, that Aristotle holds the hedonist view that the distinguishing mark of what is good, is that it is pleasant? The obvious objection to this would be that not everything pleasant is good. Aristotle makes this objection himself, at NE 1148b15 ff.: there are such things as *unnatural* pleasures. However, he also meets this objection, at 1153b8-9: τῷ ἄριστον δ' οὐδὲν κωλύει ἡδονὴν τινα εἶναι, εἰ ἔνιαι φαῦλαι ἡδοναί. The way now seems clear for the ethical hedonist view of Aristotle to go through. The good for humanity is happiness (NE 1097b21), and happiness is attained (let us say, making rough and ready use of the ἔργον argument) either in the unimpeded performance of a natural activity (NE VII) or in the perfection of such an activity (NE X). Therefore 'pleasure accompanies the

good' (τὴν εὐδαιμονίαν μεθ' ἡδονῆς, NE 1152b7), and indeed is its distinguishing mark. We can do our moral epistemology by assuming ethical hedonism.

In spite of all this, I still think that it is fairly obvious that the usual verdict of modern scholarship is right: Aristotle is not an ethical hedonist, nor indeed any sort of hedonist. In Justin Gosling's words, Aristotle's concessions to hedonism 'should be sweet to the hedonist's ears so long as he looks only at the score and does not face the music' (Gosling 1973/4).

The ethical hedonist's reading of Aristotle's moral epistemology can withstand the admission that the presence of the good is not identified by the presence of *just any* type of pleasure. He can simply refine his thesis to say that the presence of the good is identified by the presence of *the right* type of pleasure. However, what is this? What is it that all right pleasures, and no wrong pleasures, have in common which enables us to identify their presence, and hence the presence of the morally good?

Here the ethical hedonist's progress meets an insurmountable *mauvais pas*. Perhaps what he would like to find is, ideally, some particular, some more than merely formal, *quality* which all good pleasures and no bad pleasures share. That quality will be the key to everything: an Alchemists' Stone for moral epistemologists. But from this position, a more plausible reading of Aristotle's account checkmates the ethical hedonist's reading

in three moves. (1) There is *no* single, more than merely formal quality which all good pleasures and no bad pleasures share. (2) Not only that, there is only one *merely formal* quality which all good pleasures and no bad pleasures share. (3) The nature of that single merely formal quality which all good pleasures and no bad pleasures share is such as to make the ethical hedonist reading of Aristotle impossible because redundant.

To explain these moves. (1) The only *definitions* of pleasure which the NE gives are the (notoriously different) formulations found in NE VII and X. The NE VII account tells us that pleasure is the unimpeded performance of a natural activity (NE 1153a13-15); the NE X account, that pleasure 'perfects the activity', τελεῖται δὲ τὴν ἐνέργειαν ἢ ἡδονή (NE 1174b23). There is nothing in either definition which says that it applies only to good pleasures and not to bad pleasures.

Moreover, both definitions cover a remarkably wide range of experiences: *anything* that counts as an 'unimpeded performance of a natural activity', or as the 'perfection of an activity', will count as a pleasure for Aristotle. Thus he tells us that 'each sense has its own pleasure' (NE 1174b21), and that it is not the same pleasure which all creatures pursue: 'for nature and the best state neither is nor seems the same for them all' (NE 1153b29). Again, 'each person exercises his faculties concerning what he loves best: for example, the musician exercises his sense of hearing on melodies, the intellectual exercises his understanding on theoretical problems' (NE 1175a13-15): the

inference is that these very different activities involve very different pleasures, an inference which, at 1175b1, Aristotle goes on to draw. In fact, teleological speculations apart (b30-32), there seems to be nothing at all necessarily in common between the pleasures which different creatures pursue *except* the definitional requirement that all the things that they pursue should be unimpeded performances of natural activities: οὐδ' ἡδονὴν διώκουσι τὴν αὐτὴν πάντες, ἡδονὴν μέντοι πάντες (NE 1153b29). This is why I say that there is *no* single, more than merely formal quality which all good pleasures and no bad pleasures share: ὥσπερ οὖν αἱ ἐνέργειαι ἕτεραι, καὶ αἱ ἡδοναὶ (NE 1175b37).

(2) What, then, is the single merely formal quality which all good pleasures and no bad pleasures share? The answer to this is that the only thing that all good pleasures have in common is that they are the (characteristic) pleasures *of a good person* (NE 1176a17 ff.). But (3) if this is so, then the ethical hedonist reading of Aristotle's moral epistemology is no longer defensible. For the ethical hedonist began by saying that 'the presence of the good was identified by the presence of pleasure'. Under pressure, this was amended to: 'the presence of the good is identified by the presence of *the right type* of pleasure'. Under further pressure, it was conceded that the right type of pleasure, the presence of which is diagnostic of the presence of the morally good, was not qualitatively but only formally definable. It now appears that the correct formal definition of the type of pleasure involved is 'the characteristic pleasure(s)

of the good person'. (Hence Aristotle's question 'What is the best pleasure?' (NE 1176a25) is just another way of asking 'What is the best *activity*?', as NE 1177a13 ff. shows.) But then the ethical hedonist thesis, as refined, now says that 'the presence of the good is identified by the presence of the pleasure(s) characteristic of the good person'. It seems that, once again, we can only make use of the ethical hedonist criterion of good pleasures if we have a prior understanding of a quite different notion, that of the good person. For 'good pleasures' can only be defined with respect to this good person in whom they occur.

2g. Aristotelian Epistemology and the Man of Virtue

We have seen that attempts to base Aristotelian moral epistemology on the notions of function, or the mean, or the intuition of indefeasible moral rules, or on the intuition of moral principle or moral particulars, or on pleasure, are all bound to be unsuccessful for the same simple reason: all these notions imply a prior notion, that of the *expert in applying or defining them*. It is an important part of the argument for basing our moral epistemology on this notion that we should see that other attempts to find such a basis fail; and that we should see that there is a quite general reason why they fail, namely that they all refer us back to this prior notion of the expert. I now examine that notion.

We saw (§2c) that the expert in applying moral rules, where

the formulation of these is appropriate, is ὁ νομοθέτης (NE 1137b20 ff.); and νομοθεσία is a subspecies of φρόνησις (1141b34), which itself is part of intellectual virtue (1139b14 ff.). (§2a) The expert in applying the notion of function is the person who is good at defining τεχναῖ, which, again, is an exercise of intellectual virtue; and the person who is good at defining τεχναῖ also turns out to be the expert in applying Aristotelian moral principles (§2d)– without which expertise there can be no φρόνησις (§2e). Again (§2b), who is the expert in applying the concept of the mean? Aristotle's geometrical analogy is suggestive: ἐν ἐκάστῳ γὰρ τὸ μέσον λαβεῖν ἔργον, οἷον κύκλου τὸ μέσον, οὐ πάντος ἀλλὰ τοῦ εἰδότος (NE 1109a25–26). Who, finally, is the only possible measure of what counts as a good or bad pleasure? The answer (§2f) is ὁ ἀγαθός. As shown by the very passage which makes it clear that an ethical hedonist reading of Aristotle is a non-starter, the expertises of applying these various methods of finding what is good converge in the overall expertise in moral epistemology of Aristotle's 'man of virtue' (and I use the word 'man' advisedly):

δοκεῖ δ' ἐν ἅπασιν τοῖς τοιούτοις εἶναι τὸ φαινόμενον τῷ σπουδαίῳ... καὶ ἔστιν ἐκάστου μέτρον ἢ ἀρετὴ καὶ ἀγαθός ἢ τοιοῦτος, καὶ ἡδοναὶ εἶεν ἂν αἰ τούτῳ φαινόμεναι καὶ ἡδέα οἷς οὗτος χαίρει. (NE 1176b17–19)

At 1109a25–26 (above) Aristotle makes the general claim that the way to understand any particular τέχνη is to ask the expert in that τέχνη: cp. 1094b29, ἕκαστος δὲ κρίνει καλῶς ἃ γινώσκει,

καὶ τοῦτων ἐστὶν ἀγαθὸς κριτῆς. In pursuing these extensions of the ἔργον argument, it becomes clear that the way to understand virtue, *qua* hierarchy of τεχναὶ, is to ask the expert in virtue.

One has, of course, to speak guardedly when arguing that virtue, in Aristotle's conception, is in important respects like a τέχνη. There are important differences too. (i) Virtue is concerned with πράξεις, τέχνη with ποιήσεις. (ii) Voluntary errors in πράξεις are worse than involuntary ones, but in τέχνη the opposite is true (NE 1140b23-24). (iii) Excellence in πράξεις is mainly to do with habituation, excellence in a τέχνη with knowledge (NE 1105b2-4). (iv) A craftsman is technically excellent if and only if his production is technically excellent; but a moral agent is not virtuous if and only if his action is virtuous (1105a27-30). As it happens Aristotle does sometimes compare πράξεις to loose kinds of τέχνη, like medicine and helmsmanship (NE 1104a8-11); but the only similarities on which my argument here depends are these two. (i) Excellence in πράξεις can, like excellence in a τέχνη, be explicated by means of the ἔργον argument. And, therefore, (ii) excellence in πράξεις can, like excellence in a τέχνη, be understood as an expertise.

Granted these riders, my claim is that the expert in virtue is *the man of virtue*:

περὶ πάντα μὲν ταῦτα [τοῦ καλοῦ, συμφέροντος, ἡδέος] ὁ ἀγαθὸς κατορθωτικός ἐστὶν ὁ δὲ κακὸς ἀμαρτητικός. (NE 1104b32)

τοῦτο δ' [sc. τὸ τέλος καὶ τὸ ἄριστον] εἰ μὴ τῷ ἀγαθῷ οὐ φαίνεται. (NE 1144a32)

τοῦ γὰρ φρονίμου μάλιστα τοῦτ' ἔργον εἶναί φαμεν, τὸ εὖ βουλεύεσθαι. (NE 1141b10; cp. 1142b31)

What does this claim mean? What it means is that *expertise in virtue* is not distinct from *excellence of character*: the man of virtue's expertise, what makes him the standard of virtue, is his excellence of character.

However (1105b1 ff.), this does not mean that it is sufficient for me to have moral knowledge that I should ask the man of virtue about what is good. To know, in the true sense of 'know', what is good, I must myself *become* the man of virtue (which incidentally makes virtue, or full virtue, impossible for women on Aristotle's account). Merely to know, in an abstract sense, which ethical propositions are true and which false, is meaningless for Aristotle and does not count as genuine moral knowledge: τὸ δὲ λέγειν τοὺς λόγους τοὺς ἀπὸ τῆς ἐπιστήμης οὐδὲν σημεῖον (NE 1147a18–23). Nor, indeed, is it sufficient for me to have moral knowledge that I should act in a virtuous way, that is to say in the way the man of virtue acts or would act. Doing the kind of things the man of virtue does is necessary for becoming a man of virtue; but it is not sufficient unless I do the kind of things the man of virtue does *in the way that the man of virtue does them*:

τὰ μὲν οὖν πράγματα δίκαια καὶ σώφρονα λέγεται, ὅταν ᾗ τοιαῦτα οἷα ἂν ὁ δίκαιος ἢ ὁ σώφρων πράξειεν· δίκαιος δὲ καὶ σώφρων ἐστὶν οὐχ ὁ ταῦτα πράττων, ἀλλὰ καὶ ὁ οὕτω πράττων ὡς οἱ δίκαιοι καὶ οἱ σώφρονες πράττουσιν. (NE 1105b5-9)

This is why Aristotle reminds us that the objective, in a study of ethics, is a *practical* one:

ἡ παροῦσα πραγματεία οὐ θεωρίας ἕνεκά ἐστιν ὥσπερ αἱ ἄλλαι (οὐ γὰρ ἔν' εἰδῶμεν τί ἐστὶν ἡ ἀρετὴ σκεπτόμεθα, ἀλλ' ἔν' ἀγαθοὶ γενώμεθα, ἐπεὶ οὐδὲν ἂν ᾗν ὄφελος αὐτῆς). (NE 1103b26-29)

ἀλλ' οἱ πολλοὶ ταῦτα μὲν οὐ πράττουσιν, ἐπὶ δὲ τὸν λόγον καταφεύγοντες οἴονται φιλοσοφεῖν καὶ οὕτως ἔσεσθαι σπουδαῖοι... (NE 1105b12-15)

Aristotle believes that there is no neutral and objective definition or description of virtue available which is equally intelligible and acceptable to all comers, irrespective of their attitudes and characters. There *is* an objective understanding of virtue; but not a neutral one. The whole point about conceiving virtue as a τέχνη is that, if that conception is right, then, just as with the other τεχναὶ, it is the opinions of the experts that are conclusive. If you want to have the right opinions about what human excellence consists in, you have no option but to engage in the τέχνη of virtue. But this means being committed to becoming *good*: to becoming a man of virtue.

What are we to make of this doctrine? At first sight it might just seem a mistake to say that what is good, is good *because* it is what the good person thinks is good. Surely (it might be replied), the truth is rather that what the good person thinks is good, is good *because the good person is good*? Surely we should define the good person in terms of the good, and not the good in terms of the good person? But Aristotle clearly does mean to give *the good person*, and not *the good*, definitional priority. His reason for doing this is explained at NE 1113a32-34:

καθ' ἐκάστην γὰρ ἕξις ἰδιά ἐστι καλὰ καὶ ἡδέα, καὶ διαφέρει πλεῖστον ἴσως ὁ σπουδαῖος τῷ τἀληθὲς ἐν ἐκάστοις ὄραν, ὥσπερ κανὼν καὶ μέτρον αὐτῶν ὢν.

The reason why the good is to be defined in terms of the good person, and not *vice versa*, is because τἀγαθὸν ἰσαχῶς λέγεται τῷ ὄντι (NE 1096a24). Aristotle writes against (*inter alios*) Platonists, and so rejects the idea that there is a single, abstract thing called The Good, in terms of which all human goods are to be commensurated. Yet this does not mean that he rejects all possibility of commensurating goods. There is for him no single sense of Good, as the Platonists believe, in terms of which all other uses of 'good' are to be understood. (Cf., for example, the notorious division between action and contemplation noted above: this seems to be a clear case where there is not *meant* to be an easy way of commensurating what I have called τέλος and meta-τέλος.) But, on the other hand, it is not true that there is no chance of relating the different uses of 'good'

to each other: they do not seem to have appeared by *accident* (οὐ γὰρ ἔοικε τοῖς γε ἀπὸ τύχης ὁμωνύμοις, NE 1096b26). Aristotle's suggestion is that the locus of commensuration between these different senses is the *life of the good person*. Different supposed goods, if they are truly good, will find their places in the life of the man of virtue. Hence the way to find what is good is not to search for a definition of 'good' in the abstract, but to look for a concrete example of a *good life*. Aristotle would not have said, with Protagoras, that 'Man is the measure of all things'; but he might have said that '*The man of virtue* is the measure of all things'.

The obvious response to this claim is: But *who* is the man of virtue? It may seem that very little has been established in what I take to be Aristotle's moral epistemology, apart from the principle that the good is to be defined in terms of the good person and not *vice versa*. But this principle is unenlightening, unless we already know who the man of virtue is. In fact it seems that the same circularity applies which I noted before. Just as (I claimed) we need the prior notion of expertise in defining notions like 'function' and 'the mean' before we can make use of them, so we need the prior notion of expertise in defining the notion of the man of virtue before we can make use of the notion of the man of virtue.

In response to this objection, I think two points can be made. First, if the request for a definition of the man of virtue is an attempt to take us back from true ethics, done ἔν' ἀγαθοῖ

γενώμεθα, to the method of inquiry of an ordinary ἐπιστήμη, done θεωρίας ἕνεκα (1103b26-29), then Aristotle would certainly resist this. Just as for Aristotle there is no definition of the concept of virtue which is equally available to all, regardless of their own excellence of character or otherwise, so likewise it is indeed the case that no definition of what it means to be a man of virtue could be neutral in this way. But if it is (as Aristotle clearly believes) *true* that there can be no definition of the man of virtue which is equally accessible to idle intellectual curiosity and to sincere moral aspiration, then it is not a *failing* of his theory that it does not attempt, like so many other theories, to provide what *cannot* be provided. Aristotle *wants* us to see, first, that rules of thumb for finding the good (such as his own concepts of function and the mean) are useless unless we have a prior notion of expertise in their application; and, second, that the very notion of moral expertise is not separable from the notion of being good. To recognise these points in response to his theory, then, is not to refute him.

Second, following on from this, Aristotle does in any case have an explicit description, and an implicit definition, of the man of virtue. The explicit description is given by the portrait of goodness of character which we find in NE III.6-VI.13. On the basis of Aristotle's important dictum (1103b26) that ethics is done οὐ θεωρίας ἕνεκα, ἀλλ' εἰν' ἀγαθοῖ γενώμεθα, we may add to the details of character given by this description that Aristotle's implicit definition of the man of virtue is this: he

is the person who, like Aristotle himself in his own ethical writings, is engaged in a morally serious attempt to define the man of virtue.

So it is not quite true, in Martha Nussbaum's words, that 'Aristotle pointedly fails to endorse an art of measuring for ethics' (Nussbaum 1986, p.290). There is a measure, the Man of Virtue himself: ἔστιν ἐκάστου μέτρον ἢ ἀρετὴ καὶ ἀγαθός ὃ τοιοῦτος (1176b18). But the paradox is that, for me to be skilled in the use of this measure, it is necessary that I myself should become the man of virtue.

Hence the definition of the man of virtue which Aristotle can offer us is both provisional and (in a sense) circular in character. But neither its provisionality nor its circularity renders the definition negligible. On the contrary, these features are its strengths. The definition of the man of virtue has to be *provisional*, because virtue is in an important sense a τέχνη (though it is not, of course, just like other τεχναὶ). A τέχνη is, as 52c suggested, an activity one crucial part of which is precisely to continue refining its own definition. And it has to be *circular*, or at any rate to proceed in a spiral, because as we have seen there is no way of understanding the enterprise of human excellence without being committed to that enterprise oneself.

Both these points come out in the technique of moral (and, indeed, other) epistemology which is implied by and visible in

Aristotle's writings. Aristotle inherits from his philosophical forebears, especially Plato, the crucial epistemological contrast between appearance and reality. Appearance, for Plato and Parmenides, was absolutely deceptive, and reality was absolutely dependable. Aristotle, however, transformed this antithesis into a quite different contrast, in which both terms of the antithesis have something important to contribute to the growth of knowledge. Aristotle's contrast is between the commonly held opinions of humanity, and the views of the wise.

The 'commonly held opinions of humanity' are given in the ἔνδοξα (NE 1145b5-7), the φαινόμενα (1145b3), the λεγόμενα (1098b11, 1145b20), whatever δοκεῖ (1094a2, 1142a1, 1145b8), and the proverbs which Aristotle quotes (e.g. 1155a33 ff.). The 'views of the wise' are those of οἱ ἔνδοξοι ἄνδρες (NE 1098b28), of ὁ σπουδαῖος (1099a23), ὁ ἀγαθός (1176b19), οἱ φρόνιμοι (1095b28), ὁ νοῦν ἔχων (1112a21); of the poets, philosophers and sages whom Aristotle quotes (1152a23 (Anaxandrides), a32 (Euenus); 1155b2 (Euripides), b4 (Heraclitus), b7 (Empedocles)).

The first point to note about the two sides of this contrast is that Aristotle does *not*, like Plato, treat one side as altogether bad, and the other as altogether good. We see them working together at NE 1098b27-30:

τούτων δὲ τὰ μὲν πολλοὶ καὶ παλαιοὶ λέγουσιν, τὰ δὲ ὀλίγοι καὶ ἔνδοξοι ἄνδρες· οὐδετέρους δὲ τούτων εὐλογον διαμαρτάνειν τοῖς ὄλοις, ἀλλ' ἔν γέ τι ἢ καὶ τὰ πλεῖστα κατορθοῦν. (NE 1098b27-30)

In many cases, indeed, the views of the many correct those of the wise. So Speusippus is upbraided for arguing, highly counter intuitively as Aristotle thinks (NE 1094a1-2), that not all creatures seek the good:

ἀ γὰρ πᾶσι δοκεῖ, ταῦτ' εἶναί φασιν· ὁ δ' ἀναιρῶν ταύτην τὴν πίστιν οὐ πάνυ πιστότερα ἔρεῖ. (NE 1173a2)

It is made clear elsewhere that this very sweeping claim is not, in fact, made by Aristotle without qualification. For the views of the wise can also correct those of the many. οἱ πολλοὶ can sometimes be a term of opprobrium for Aristotle, as it standardly is for Plato:

ἀλλ' οἱ πολλοὶ ταῦτα μὲν οὐ πράττουσιν, ἐπὶ δὲ τὸν λόγον καταφεύγοντες οἴονται φιλοσοφεῖν καὶ οὕτως ἔσεσθαι σπουδαῖοι... (NE 1105b12-15; cp. 1099a12)

The views of the wise are just those refined and chastened views which correct the views of the many. Still, these views are themselves to be subjected to further scrutiny concerning their relationship to the views of the many. Philosophy has a duty to *improve on* the common opinions of humanity, for those opinions are often in crude conflict. It also has a duty to *stay loyal to* those opinions, for they represent the place where philosophy is best advised to begin its investigations. This may give a context to the sweeping claim of 1173a2. Aristotle's point, in context, cannot be the claim that the ἔνδοξα are always true. It is more

likely that, on balance, he would have said something like J.L. Austin:

'Certainly ordinary language has no claim to be the last word... in principle, it can everywhere be supplemented and improved upon and superseded. Only remember, it *is* the *first* word.'

(Austin 1956/7)

Aristotle thinks that glibly commonsensical philosophy, such as Heraclitus accused Homer and Hesiod of purveying (DK B56, B57), is bad philosophy; but he thinks that wildly counter intuitive philosophy, philosophy which begins for example with universal doubt, is bad philosophy too. (Note that this does *not* mean that Aristotle would simply shrug off radical sceptical questions; although it does mean that he would criticise the form as well as the content of the inquiry that raised them.) Hence, for Aristotle, a common opinion which is plainly illogical or fatuous is not, in spite of 1173a2, immune from criticism simply because it is widely held. And a philosophical conclusion which ἀμφισβητεῖ τοῖς φαινόμενοις ἐναργῶς, like Socrates' about akrasia (1145b28), is not to be tolerated either, however venerable its provenance. The relationship between the two sides of the contrast is *dialectical*, in the sense that there is a continuous mutual correction and refinement of each other by the two sides of the dialectic.

Where do Aristotle's own arguments stand in relation to that dialectic? The answer to this comes when we consider an objection to seeing the relationship between the views of οἱ πολλοὶ καὶ οἱ χαρίεντες (1095a18) as I have just suggested. It might be pointed out that Aristotle seems to see the views of οἱ πολλοὶ καὶ οἱ χαρίεντες rather as one loose category than as two specific ones which are in opposition. So, above, I quote proverbs on the one side of my contrast, and Anaxandrides, Euenus, Euripides, etc. on the other, from *the same list* of δοξαὶ at 1155a33 ff.. But what this demonstrates is not that Aristotle does not see a dialectical relationship between the views of οἱ πολλοὶ and οἱ χαρίεντες. Rather, it shows that that dialectical contrast has both a historical and a current application.

In the historical sense, someone is a σοφός relative to any set of common opinions when she attempts, and at least to some degree proves able, to put them in coherent order. The famous σοφοὶ καὶ χαρίεντες of the past are exactly those persons who took generally received beliefs and systematised them where they did not fit together. But what count in this historical perspective as the views of οἱ πολλοὶ and those of οἱ χαρίεντες, are *all*, from the current perspective, 'commonly held opinions'. Relative to these it is our own current efforts at clarification, definition or refutation which will count, if successful, as the views of οἱ σοφοί. They will count as such, of course, for just as long as they resist *further* refinement or refutation; from then on, they will only be the views of οἱ σοφοί in the historical perspective. Thus Aristotle's verdicts stand to all

the views which he considers, and refines, refutes or reaffirms, in the same relationship which some, but not all, of those views once stood to others of those views. And likewise, present day attempts to extend, expose or explicate Aristotle's project also stand in that same relationship to Aristotle's views, and to those to which his views once stood in that relationship.

Thus Aristotle's method is to look for those views which *either* are 'most widely current' (the many), *or* 'seem to have some logic to them' (the wise) (1095a28-30); and for a way of reconciling them where they clash. When his method leads him (1095a14-30) to see that the popular account of happiness differs from the Platonists' account, and that *both* views need refinement, this clash is a good example of a typical starting point for his philosophy. It is in this context that we may recall his famous dictum at NE 1145b2-8: that the philosopher should begin with the apparent views and their difficulties, and establish as many as possible of the ἔνδοξα. If not all of them can be established, then the preference is to be for τὰ πλείστα (the views of the many) καὶ κυριώτατα (those of the wise):

ἐὰν γὰρ λύηται τε τὰ δυσχερῆ καὶ καταλείπηται τὰ ἔνδοξα, δεδειγμένον ἂν εἴη ἰκανῶς... τούτων δὲ τὰ μὲν ἀνελεῖν δεῖ τὰ δὲ καταλιπεῖν· ἢ γὰρ λύσις τῆς ἀπορίας εὔρεσις ἐστίν. (1145b6-7 and 1146b6-7)

To describe this dialectic is to describe the functioning of a τέχνη, the very special τέχνη of virtue, within a *philosophical*

tradition. Aristotle believes that moral epistemology is itself part of the practice of a τέχνη: a skill or craft the practice of which is under constant revision and refinement. This means that we see ourselves as part of a *history*: our work within the craft has its validity from the fact that it is part of that craft. It also means a commitment, clearly visible in Aristotle's attitudes, to epistemological conservatism: where possible, we take on board what has already been settled. For ὁ δ' ἀναιρῶν ταύτην τὴν πίστιν οὐ πάνυ πιστότερα ἔρεϊ. Further, it means philosophical humility: our insights, however thrilling, are indeed to be taken as truth, but also as provisional truth. There is nothing to guarantee that our current conception of the man of virtue will not be radically and rightly changed in the near future, and we should be always be open to that possibility; but until that happens, our current conception is the best we have, and we have no reason whatever not to act on it.

which

also Moral epistemology turns out to be a τέχνη which is part of the τέχνη of virtue. What does general epistemology turn out to be? The answer is that general epistemology is a τέχνη too, which is conducted in something like the sort of way just suggested for the τέχνη of virtue. There will be an expertise in general epistemology, the ability to have αἴσθησις (in a more physiologically 'basic' sense of αἴσθησις than that relevant to φρόνησις) of what is really the case. This expertise will be defined by reference to the notion of an expert in general epistemology.

offers us

What is the connection between this kind of expertise and *overall* expertise, that is to say the excellence of character proper to a man of virtue? The answer to *this* is that to be an expert in general epistemology is also part, though a less basic part, of that same excellence of character, just as any other expertise in a τέχνη is part of excellence of character (1113a33: καὶ διαφέρει πλεῖστον ἴσως ὁ σπουδαῖος τῷ τἀληθές ἐν ἐκάστοις ὁρᾶν). To be an expert in moral epistemology is to be a man of virtue, *sans phrase*. To be an expert in general epistemology is to be a man of virtue *with respect to general epistemology*.

Thus, again, I am blamed for all principle ignorance, because (as Aristotle thinks) *all* principle ignorance has its ἀρχὴ in me; but only some particular ignorance has its ἀρχὴ in me (1135b18-19), and it is only for this that I am blamed. I am blamed for this latter because one of the kinds of habituation to which I *ought* to submit myself is habituation towards correct and accurate awareness of the physical world. Insofar as I can so habituate myself, it is my responsibility to do so. Insofar as I do so habituate myself, I become an ἀγαθός. In this respect, the τέχνη of general epistemology is not importantly different from any other Aristotelian τέχνη except the special case of moral epistemology and virtue: for *that* τέχνη is the one of which all τέχναί, properly and improperly so called, are parts.

This completes what I have to say about the positive accounts of knowledge and origination of an impulse to act which Aristotle offers us as corollaries of his negative conditions of voluntary

action that such action should (i) not be compelled, and (ii) not be done in ignorance. What, then, of the positive account of rationality which I say Aristotle offers us as a corollary of his third negative condition of voluntary action, that such action should not be done irrationally? To that question I turn in Chapter Three.

Chapter 3

Proairesis and Practical Reasoning

1. προαίρεσις
2. τὸ ἐκούσιον, Action on προαίρεσις and the Difference between them
3. βούλευσις
4. Practical Reasoning and the 'Practical Syllogism'
5. Conclusion

'Quand je délibère, les jeux sont faits.'

(Sartre 1945)

ποῖα δ' ἀντὶ ποίων αἰρετέον, οὐ ῥάδιον ἀποδοῦναι· πολλαὶ γὰρ
διαφοραὶ εἰσιν ἐν τοῖς καθ' ἕκαστα.

(Aristotle, NE 1110b7)

In Ch.2 I discussed the positive accounts of unhindered appetite and of knowledge which Aristotle offers us as corollaries to his negative requirements of voluntary action, that it should be behaviour which is neither compelled nor done in ignorance. Since the third negative requirement of voluntary action was that it should not be behaviour which is *irrational*,

in a sense I have defined, it is clearly time to examine Aristotle's positive account of the rationality of action.

It may be clear why I have left this third positive account till last. The rationality of an action consists in the *combining* in that action of an unhindered appetite with relevant knowledge. Therefore it is plain that we cannot discuss Aristotle's account of the rationality of action without first being clear about his accounts of appetite and hindrance, and of knowledge and relevance. Moreover, we cannot discuss this third part of Aristotle's positive account of voluntary action without constantly presupposing and referring to the first and second parts of his account. It is impossible to deal with his doctrine of rationality in isolation.

Aristotle's positive account of rationality is a treatment of *how* relevant knowledge and unhindered appetite combine to be the cause of a voluntary action. Examining what he has to say about this leads me to consider, first προαίρεσις (§1-2), then βούλευσις (§3), and third, Aristotle's doctrine of the 'practical syllogism' (§4).

1. προαίρεσις

(d) What is προαίρεσις? Aristotle discusses it in detail in two passages of the NE, III.2 and 1139a21-b13. I will concentrate here on III.2, as this actually purports to give us a definition

of προαίρεσις. Aristotle works his way towards this definition by giving us a list of things which προαίρεσις is *not*. Aristotle aims to establish, on the one hand, that προαίρεσις is not simply an ὄρεξις, an appetite; and, on the other, that it is not simply a δόξα (opinion or belief) either (NE 1111b11-13). This, I take it, is because Aristotle wants to argue that προαίρεσις involves elements of both appetite and belief. That is certainly the upshot of the VI.2 discussion: διὸ ἢ ὄρεκτικὸς νοῦς ἢ προαίρεσις ἢ ὄρεξις διανοητική (NE 1139b5).

For Aristotle as for the Plato of the *Republic*, there are three kinds of ὄρεξις, corresponding to the three parts of the soul: ἐπιθυμία (desire), θυμὸς (spirit), and βουλήσις (wish). No one of these ὄρεξεις is identifiable with προαίρεσις, though the third might have seemed the most plausible candidate to Aristotle's contemporaries. ἐπιθυμία is disqualified from being identifiable with προαίρεσις for these reasons:

(a) 'προαίρεσις is not shared with irrational creatures (τῶν ἀλόγων), but ἐπιθυμία and θυμὸς are.' (1111b13)

(b) 'The ἀκρατής acts on desire (ἐπιθυμῶν), but not on προαίρεσις; whereas the self disciplined person acts on προαίρεσις, but not on desire.' (b14-15)

(c) 'προαίρεσις can oppose desire, but *desire* cannot oppose desire.' (b16)

(d) 'Desire is concerned with the pleasant and the painful, προαίρεσις with neither.' (b17)

θυμὸς is disqualified too (b19), although Aristotle gives us little explicit argument for this claim apart from (a) above. Possibly (a)-(d) are all meant to apply to θυμὸς too, *mutatis mutandis*.

Aristotle gives two reasons why βούλησις cannot be equated with προαίρεσις, καίπερ σύνεγγυς φαινόμενον (1111b20-34):

(e) There can be βούλησις for things which are either impossible *tout court*, or impossible *for me*; there cannot be προαίρεσις of either (b21-26; cp. 1139b7-13).

(f) βούλησις is of the end (τοῦ τέλους), προαίρεσις of the means (τῶν πρὸς τὸ τέλος) (b27-29).

Given the success of these six arguments, and given Aristotle's classification of ὄρεξις into three sorts, it follows that προαίρεσις is no kind of ὄρεξις. The mere presence of an appetition, of whatever kind, is no guarantee of the presence of a προαίρεσις.

(*Contra* Charles 1984, p.58, the point of NE 1139a23 cannot be that προαίρεσις is an ὄρεξις; unless Aristotle there simply contradicts the arguments of NE 1111b11-30 which I have just outlined. If there are only three sorts of ὄρεξις, and προαίρεσις is none of those three, then it seems clear that προαίρεσις cannot be any ὄρεξις at all. The point of NE 1139a23 (cf. 1113a11) is, rather, the point amplified by 1139b5: διὸ ἢ ὄρεκτικὸς νοῦς ἢ προαίρεσις ἢ ὄρεξις διανοητική. προαίρεσις is

not simply a process of thought, διάνοια; but it is not simply a process of appetition, ὄρεξις, either. προαίρεσις is the case where they *combine*, in a product which is just as much ὄρεξις as it is διάνοια and which, properly speaking, is more complex than either alone.)

For Aristotle then argues that προαίρεσις is not simply δόξα either. He begins this discussion by saying that ἡ μὲν γὰρ δόξα δοκεῖ περὶ πάντα εἶναι (1111b31). It seems that δόξα here means 'belief' in its widest sense: not that sense in which 'belief' is opposed to knowledge, as at 1139b17 and 1145b32 ff., but a sense which includes both knowledge *and* (mere) belief, as suggested by τοῦ δοξαστικοῦ at 1140b27. That wide sense certainly seems to be the one employed at a11, where Aristotle says that δόξα precedes and accompanies προαίρεσις- having just remarked (a7) that προαίρεσις involves certain knowledge, not hazy belief. (For more on 1112a7-8, see below.)

The suggestion, then, is that the mere presence of a belief, of whatever kind, is no guarantee of the presence of a προαίρεσις either. Why not?

(g) (cf. (e)) Belief can concern any matter at all, not just τὰ ἐφ' ἡμῖν (1111b33; cp. 1139a36 ff.).

(h) Belief is qualified as either true or false; but προαίρεσις, rather (μᾶλλον), as right or wrong (b34-35; cp. 1139a21-22).

(g) and (h) are offered to show that προαίρεσις is not

identical to *any kind* of belief. Aristotle then continues:

ἀλλ' οὐδέ τινα· τῷ γὰρ προαιρεῖσθαι τὰγαθὰ ἢ τὰ κακὰ ποιοῦ
τινές ἐσμεν, τῷ δὲ δοξάζειν οὐ. (NE 1112a1-3)

This sentence introduces six more specific arguments (i-n) for the claim that προαίρεσις is not identical to τινα δόξα, 'to some particular belief', by which Aristotle seems to mean *moral* belief. If (g) and (h) succeed, then (i-n) are strictly unnecessary; but Aristotle obviously thinks the point worth stressing.

(i) προαίρεσις makes our characters good or bad- not moral belief (a1-3; cp. 1139a33-34).

(j) (cf. (g)) προαίρεσις is practical, moral δόξα theoretical (a3-5).

(k) (cf. (h)) Moral belief is qualified as either true or false (τῷ ἀληθῶς),
προαίρεσις as either right or wrong (τῷ ὀρθῶς) (a6-7).

(l) What we *believe* (δοξάζομεν) to be right, we do not *know for sure* (οὐ πανύ ἴσμεν) to be right. But action on προαίρεσις is action on knowledge of what is right, not on (mere) belief (a7-8).

(m) I can have good 'abstract' moral beliefs without being a good person (a9-10; cp. 1147a22).

(n) That moral belief is a necessary condition of προαίρεσις does not show that moral belief *is* προαίρεσις (a11-13; cp. 1139a33).

(a-n) are, as I say, presented by Aristotle as negative arguments. But plainly they also give us a good idea of the cards in Aristotle's own hand. Given a clarification of (f) and the exclusion of (1), the combination of these arguments with Aristotle's positive remarks gives us a fair account of his doctrine of προαίρεσις.

(1) should be excluded from consideration if, as I have argued, the whole point of NE III.2 is to show that προαίρεσις involves the combination of unhindered appetite and relevant knowledge, and hence δόξα here means 'belief' in its widest sense, not that sense in which 'belief' is opposed to knowledge. Of course it is true that action on προαίρεσις is action on knowledge, not on (mere) belief. Action on belief which is not also knowledge is action in ignorance, and hence not voluntary action, and hence (*a fortiori*) not action on προαίρεσις either. But this is a point against δόξα as opposed to knowledge, not against δόξα in the wide sense which *includes* knowledge. Aristotle's aim, I am suggesting, is to show that προαίρεσις must mean the combination of unhindered appetite and relevant knowledge, because either alone is insufficient for προαίρεσις. Then attacking the idea that προαίρεσις could be equivalent to δόξα in the narrow sense, as (1) does, is beside the point.

The clarification of (f) is this. 1111b27 need not mean that the setting up of a τέλος is an altogether non-rational process. Aristotle's pronouncement that there is no λόγος of definitions (1142a26) does not mean that the setting up of a ὅρος is an

altogether non-rational process. It means rather that, *if* the setting up of a ὄρος is a rational process (as, we are told elsewhere, it *is*), that rational process is not, strictly, λόγος (which here means συλλογισμός). Likewise, Aristotle's pronouncement here that there is no προαίρεσις of τέλη does not mean that the setting up of a τέλος is an altogether non-rational process. It means rather that, *if* the setting up of a τέλος is a rational process (as, we are told elsewhere, it *is*), that rational process is not, strictly, προαίρεσις.

However, there are two senses of both λόγος/ συλλογισμός and προαίρεσις, broad and narrow. This leads to ambiguities concerning προαίρεσις and theoretical reasoning which I believe to be in exact parallel. Aristotle sometimes calls ἐπαγωγή a kind of συλλογισμός (APr 68b30) (the broad sense), and sometimes opposes ἐπαγωγή to συλλογισμός (the narrow sense) (68b13). Notoriously, he also describes προαίρεσις as being (i) of means only and not of ends (1111b27), (ii) of ends (1110b32, 1144a20-23), and (iii) of both means and ends (1145a4-6). There is a solution to this. Analogously to λόγος/ συλλογισμός, προαίρεσις has a narrow sense, in which it cannot include the rational processes involved in the setting up of a τέλος; but also a broad sense, in which it can.

Given these qualifications, we may proceed to outline what Aristotle tells us about προαίρεσις. (a) προαίρεσις is something distinctive to mature humans, unlike τὸ ἐκούσιον (1111b8-9). (b) It is limited not simply to mature humans, but more specifically

to self disciplined mature humans, such as those who act not suddenly (1110b10) but 'from a settled disposition of character' (1104a35). (c and d) προαίρεσις is concerned with non-abstract (m), character forming (i) beliefs (*inter alia*, n) about what is morally right (τῷ ὀρθῶς, k) in practical affairs (h, j). It can overrule any desire in a way that no desire can (c). The scope of someone's προαίρεσις is ὅσα οἴεται γενέσθαι ἄν δι' αὐτοῦ (1111b26). Thus there is no προαίρεσις about (e) impossibilities, (g) matters which are not ἐφ' ἡμῖν, or (j) theoretical questions.

These remarks establish that προαίρεσις is not simply either a belief nor an appetite. I say that this is argued to pave the way for the conclusion that προαίρεσις involves the combination of belief and appetite. So what comes next in the NE? This definition of προαίρεσις:

τί οὖν ἢ ποῖόν τί ἐστίν [προαίρεσις], ἐπειδὴ τῶν εἰρημένων [sc. ἐπιθυμία, θυμὸς, βουλήσις, δόξα: 1111b11-13] οὐθὲν; ἐκούσιον μὲν δὴ φαίνεται, τὸ δ' ἐκούσιον οὐ πᾶν προαιρετόν. ἀλλ' ὄρα γε τὸ προβεβουλευμένον; ἢ γὰρ προαίρεσις μετὰ λόγου καὶ διανοίας. ὑποσημαίνειν δ' ἔοικε καὶ τοῦνομα ὡς ἄν πρὸ ἐτέρων αἰρετόν. (NE 1112a13-17)

Aristotle, we read here, defines προαίρεσις as τὸ προβεβουλευμένον. Does πρὸ here mean 'before in time', or 'before in preference' as it does at EE 1225b8 (cp. MM

1189a13)? Joachim 1955 (p.100-101), Gauthier-Jolif 1970 (p.198), and, it appears, Kenny 1979 would argue that $\pi\rho\acute{o}$ here is temporal. A Rylean or Wittgensteinian commentator might want to argue, as might Ross (1925) and Burnet (1900), that, on the contrary, $\pi\rho\acute{o}$ is ^{non-}temporal. Hardie (1968, p.164) thinks that $\pi\rho\acute{o}$ in $\pi\rho\omicron\beta\epsilon\beta\omicron\upsilon\lambda\epsilon\upsilon\mu\acute{\epsilon}\nu\omicron\nu$ is temporal, but $\pi\rho\acute{o}$ in $\pi\rho\acute{o}$ $\acute{\epsilon}\tau\acute{\epsilon}\rho\omega\nu$ is not.

However, Hardie's distinction between $\pi\rho\acute{o}$ - and $\pi\rho\acute{o}$ seems dubious when one looks closely at a16-17. Is Aristotle really 'playing with the ambiguity of the proposition [$\pi\rho\acute{o}$]' (so Joachim according to Hardie, though in fact Joachim never suggests this)? It seems clear that what Aristotle wants to say at 1112a16-17 about $\tau\acute{o}$ $\pi\rho\omicron\beta\epsilon\beta\omicron\upsilon\lambda\epsilon\upsilon\mu\acute{\epsilon}\nu\omicron\nu$, he also wants to say about $\pi\rho\omicron\acute{\alpha}\rho\epsilon\sigma\iota\varsigma$. Hence what $\pi\rho\acute{o}$ - means in $\pi\rho\omicron\beta\epsilon\beta\omicron\upsilon\lambda\epsilon\upsilon\mu\acute{\epsilon}\nu\omicron\nu$, $\pi\rho\acute{o}$ in $\pi\rho\acute{o}$ $\acute{\epsilon}\tau\acute{\epsilon}\rho\omega\nu$ $\alpha\acute{\iota}\rho\epsilon\tau\acute{o}\nu$ should also mean- despite EE 1225b8.

There are better grounds here for saying that both $\pi\rho\acute{o}$ - and $\pi\rho\acute{o}$ are either temporal or not than for distinguishing between them as Hardie suggests. And there is further evidence for the temporal view at e.g. 1113a2-12; a passage cited by Joachim (v. 53) who writes that ' $\pi\rho\acute{o}$ is to be interpreted as having a temporal significance' (100). It is suggested that $\pi\rho\omicron\acute{\alpha}\rho\epsilon\sigma\iota\varsigma$ should be understood as a choice made *before* action or any mediate part thereof ($\pi\rho\acute{o}$ $\acute{\epsilon}\tau\acute{\epsilon}\rho\omega\nu$). $\pi\rho\omicron\acute{\alpha}\rho\epsilon\sigma\iota\varsigma$ will be a choice or decision made before acting, and with, or accompanied by, reasoning ($\mu\epsilon\tau\acute{\alpha}$ $\lambda\acute{o}\gamma\omicron\upsilon$ $\kappa\alpha\acute{\iota}$ $\delta\iota\alpha\nu\omicron\iota\acute{\alpha}\varsigma$: n.b. that $\mu\epsilon\tau\acute{\alpha}$ + genitive cannot mean 'after').

This conclusion brings us to ask: Accompanied by *what kind of* λόγος καὶ διανοία does προαίρεσις occur? The answer to that will begin to be made clear in §3, on βούλευσις. The same conclusion also puts us in a position to offer an answer to the question to which I now, at last, turn: what is the difference between action on προαίρεσις and τὸ ἐκούσιον?

2. τὸ ἐκούσιον, Action on προαίρεσις and the Difference between them

I said in Ch.1 that voluntary action (τὸ ἐκούσιον) must, for Aristotle, mean action which is (i) uncompelled, (ii) not done in ignorance, and (iii) rational. But this question then arises for my account: How do I say τὸ ἐκούσιον is different from action on προαίρεσις?

On the normal view of Aristotle, τὸ ἐκούσιον differs from action on προαίρεσις in that τὸ ἐκούσιον is any action which is (i) uncompelled and (ii) not done in ignorance. Action on προαίρεσις, by contrast, is any action which is (i) uncompelled, (ii) not done in ignorance, and (iii) rational.

Then does my claim that any voluntary action is, for Aristotle, rational, confuse τὸ ἐκούσιον and action on προαίρεσις? No: consider this conflict of texts. The definition of 'action on προαίρεσις' is 'voluntary action upon deliberation'

(NE 1112a15-17). Deliberation means practical reasoning (NE 1112b13-29). Children and animals are incapable of action on προαίρεσις, though they are capable of τὸ δ' ἑκούσιον (NE 1111b9). From these three premisses it follows that children and animals are incapable of practical reasoning. Yet animals at any rate are clearly depicted by Aristotle, in the dMA, as acting on practical reason:

ποτέον μοι, ἢ ἐπιθυμία λέγει· τοδὶ δὲ ποτόν, ἢ αἰσθησις εἶπεν ἢ ἢ φαντασία ἢ ὁ νοῦς· εὐθύς πίνει. οὕτως μὲν οὖν ἐπὶ τὸ κινεῖσθαι καὶ πράττειν τὰ ζῷα ὁρμῶσι, τῆς μὲν ἐσχάτης αἰτίας τοῦ κινεῖσθαι ὁρέξεως οὔσης... (dMA 701a32-35)

One is, of course, inclined to object to this that here Aristotle must still be talking about *human* deliberation, as he clearly was just before this passage, at 701a6-25 (a13: παντὶ βαδιστέον ἀνθρώπῳ). But in fact there is nothing in 701a32 ff. which indicates that humans alone are under discussion. On the contrary, the reference is explicitly to τὰ ζῷα. Moreover, earlier in the dMA account, Aristotle says explicitly that the behaviour of *all* animals displays just that explicability in terms of purposes which, in Ch.1, I explained was what I mean by the rationality of voluntary action:

πάντα [!] γὰρ τὰ ζῷα καὶ κινεῖ καὶ κινεῖται ἕνεκά τινος, ὥστε τοῦτ' ἔστιν αὐτοῖς πάσης τῆς κινήσεως πέρας, τὸ οὐ ἕνεκα. ὁρῶμεν δὲ τὰ κινουῦντα τὸ ζῷον διάνοιαν καὶ φαντασίαν καὶ προαίρεσιν [!!] καὶ βούλησιν καὶ ἐπιθυμίαν. (dMA 700b15-18)

Likewise NE 1118a17-23, even while belittling the idea that we could talk of other animals' having the pleasures of the senses in the same way that humans do:

οὐδ' ὁ λέων [χαίρει] τῷ φωνῷ τοῦ βοός, ἀλλὰ τῷ ἐδωδῷ· ὅτι δ' ἔγγυς ἐστί, διὰ τῆς φωνῆς ἤσθετο, καὶ χαίρειν δὴ ταύτῃ φαίνεται. ὁμοίως δ' οὐδ' ἰδὼν ἢ εὐρὼν ἔλαφρον ἢ ἄγριον αἴγα' [*Iliad* 3.24], ἀλλ' ὅτι βορὴν ἔξει.

This passage, too, clearly implies the presence of some kind of reasoning ability in beasts like lions.

How are we to account for this clash in the evidence? Must we simply say that Aristotle's attitude to animals is inconsistent between the ethical and the biological works? There certainly seems to be a strong case for saying that: compare a dismissive passage like NE 1149b32-36 with the deep interest in animals evident in the quotations above and throughout the dMA (and cp. dA 432b23-25). But we might also suggest that the theory of volition given in the dMA is meant to apply to all higher creatures, humans included.

The translators (cp. Rackham (1926), Ross (1925)) tend to assume that the scope of τὰ ἄλογα at NE 1111b13 is the same as that of τὰ ζῷα at, e.g., 1118a17-23, making both terms mean all non-human creatures. But sometimes the scope of τὰ ζῷα seems wider than this, so that it includes humans- as I am suggesting it may at dMA 770b15 ff.. By contrast the term τὰ ἄλογα at

1111b13 (its only occurrence in the NE) may well be narrower in scope, and not include the higher non-human animals. Its sense may be more like that of τῶν ἀτελεῶν at dA 433b32 than that of τὰ ζῷα at dMA 770b15.

In any case Aristotle seems much more liberal in his use of the term τὰ ζῷα in the dMA. Hence, perhaps, it is not such a glaring inconsistency for Aristotle to talk of προαίρεσις at 700b15-18. For if the scope of τὰ ζῷα includes ὁ ἄνθρωπος, then there are 'animals' which have a share in προαίρεσις. This way of reading things is partly suggested by Martha Nussbaum's work:

'When we turn to the *de Anima* and *de Motu*, we discover something that is very strange if we are used to Plato's ways of approaching the subject. Instead of Plato's moving accounts of human ethical dilemmas, we find a narrative whose leading characters are fish, birds, and insects as well as humans. Instead of what looks self-evidently important for us, we find what seems- and, we know, seemed to Aristotle's students- trivial and even disgusting. The inquiry into human action is carried out as a part of a larger inquiry into the movements of animals. Human action is very little singled out; instead we find a discussion of sweeping generality that ranges over the whole animal kingdom. It is this generality that we must seek to understand.'

(Nussbaum 1986, p.264)

Still the central question remains: what are the distinctive marks of human, as opposed to animal, voluntariness which justify Aristotle's NE doctrine that only mature humans display προαίρεσις as well as τὸ ἐκούσιον? It now appears that Aristotle offers no satisfactory answer to this. But one is clearly needed, as part of the theory of human nature given in the ethical writings. So here is my own suggestion about how to fill the gap.

My suggestion is based simply on what Aristotle himself says about action on προαίρεσις (1112a16-17): that it is to be defined as τὸ προβεβουλευμένον. This, as 53 will help clarify, means action which has been deliberated on beforehand. We can only speak of τὸ προβεβουλευμένον when there has been predeliberation; hence we can only speak of action on προαίρεσις when there has been predeliberation. If, in the case of some behaviour, there has been no predeliberation, we can speak only of τὸ ἐκούσιον or of τὸ ἄκούσιον, and not of action on προαίρεσις.

But now the problem is, not how to distinguish τὸ ἐκούσιον from action on προαίρεσις in respect of the rationality requirement, but how to distinguish τὸ ἐκούσιον, the merely voluntary which does not count as action on προαίρεσις, from τὸ ἄκούσιον in that respect. Consider some behaviour B which is not action on προαίρεσις. Although B is in fact done without predeliberation, still B may be explicable, with reference to some piece of predeliberation, as if that predeliberation had actually occurred. Then if B is so explicable, is B a case of rational action? Not necessarily, since (as I said in Ch.1) the

difference between rational action and irrational behaviour is that, in the case of rational action, it is not merely that relevant belief and unhindered appetite are present when the behaviour occurs; it is rather that the agent performs the behaviour *because of* the combination of relevant belief and unhindered appetite. But B could perfectly well be explicable with reference to some piece of predeliberation as *if* that predeliberation had actually occurred, without it being true that B occurred *because of* the combination of relevant belief and unhindered appetite.

In this case τὸ ἐκούσιον, the merely voluntary which does not count as action on προαίρεσις, must be distinct from τὸ ἀκούσιον, with respect to the rationality requirement, in this way:

τὸ ἐκούσιον is behaviour performed because of the combination in the agent of relevant belief and unhindered appetite. τὸ ἀκούσιον is any other behaviour.

And τὸ ἐκούσιον must be distinct from action on προαίρεσις, with respect to the rationality requirement, in this way:

Action on προαίρεσις is behaviour performed because of the combination *in predeliberation*, in the agent, of relevant belief and unhindered appetite. τὸ ἐκούσιον, like action on προαίρεσις, is also behaviour performed because of the combination, in the agent, of relevant belief and unhindered appetite; except that, in τὸ ἐκούσιον, that combination does

Likewise regarding actual and potential predeliberation. I might not *actually* say to myself 'Children drowning in weirs should be rescued; that child is drowning in the weir; I must jump in and save her'. It may be true that no actual thought process of predeliberation occurs, but this does not simply mean that I act without any reason or predeliberation. Rather, I act on potential predeliberation if I *would* give an account of my action by reference to such a train of thought *if I were asked*; and my behaviour is irrational if I *could* refer an inquirer to no such train of thought *if I were asked*.

I have, then, suggested how it may be that behaviour which is not action on προαίρεσις nonetheless can involve a kind of deliberation, and so be rational, as well as uncompelled and performed without ignorance, thus counting as voluntary action (as I define that); and how merely voluntary action, τὸ ἐκούσιον, differs from actually involuntary behaviour, τὸ ἀκούσιον, in respect of the rationality condition. One thing that this account makes plain is the logical dependence of the category of the merely voluntary on the category of action on προαίρεσις. Potential knowledge is defined counterfactually with reference to actual knowledge. It follows that the definition of potential knowledge presupposes that of actual knowledge. Likewise, the potential predeliberation characteristic of the merely voluntary is defined counterfactually with reference to the actual predeliberation characteristic of action on προαίρεσις. It follows that the definition of potential predeliberation, and so the definition of the merely voluntary, presupposes that of

actual pre-deliberation, and so the definition of action on προαίρεσις. It does not follow, from my rationality requirement, that the only genuinely voluntary action is action on προαίρεσις. But it does follow that the *form of explanation* for all genuinely voluntary action is that form dictated by the nature of action on προαίρεσις.

3. βούλευσις

What we want to know about next, then, is the nature of actual pre-deliberation. The natural starting point for this inquiry is Aristotle's discussion of βούλευσις, deliberation, in NE III.3. What this discussion establishes is, first, the very close connection between βούλευσις, προαίρεσις and action on προαίρεσις; and, second, that the possibility of giving a full and non-provisional account of βούλευσις (which Aristotle does not give) is dependent on the possibility of giving a full and non-provisional account of practical reasoning and 'the practical syllogism' (which Aristotle does not give either). Thus discussing βούλευσις will lead me on to the detailed discussion of practical reasoning which takes up the rest of this chapter.

First, note six close parallels between Aristotle's remarks about προαίρεσις and about βούλευσις.

(i) προαίρεσις, we heard (a), is not shared with irrational creatures; it is characteristic of sensible adult humans.

Likewise, the discussion of βούλευσις is dominated by the word βουλευόμεθα, the subject of which, if Aristotle's examples in 3.III are anything to go by, is again sensible adult humans: λεκτέον δ' ἴσως βουλευτὸν οὐχ ὑπὲρ οὗ βουλευσάιτ' ἄν τις ἡλίθιος ἢ μαινόμενος, ἀλλ' ὑπὲρ ὧν ὁ νοῦν ἔχων (NE 1112a20-21; cp. 1141b10).

(ii) There is no προαίρεσις either of what is impossible *tout court*, or of what is impossible *for me* (e). Likewise we βουλευόμεθα δὲ περὶ τῶν ἐφ' ἡμῖν καὶ πρακτῶν (1112a31).

(iii) βούλησις is of the end, προαίρεσις of the means (f). Likewise we βουλευόμεθα δ' οὐ περὶ τῶν τελῶν, ἀλλὰ περὶ τῶν πρὸς τὰ τέλη (1112b13).

(iv) προαίρεσις cannot concern any matter at all, as δόξα can, but only τὰ ἐφ' ἡμῖν (g). Likewise, (1112a21) the range of βούλευσις does not include τὰ αἰδέα or (1112b1) αἱ ἀκριβεῖς καὶ αὐτάρκεις τῶν ἐπιστημῶν (the objects of ἐπιστήμη, 1139b20); nor (1112a24) things that change with complete regularity (the objects of σοφία, 1141b1-3); nor (1112a26-27) irregular or chance changes (the objects of some sorts of τέχνη, 1140a19-20). Since (1141b2) σοφία = νοῦς + ἐπιστήμη, it would seem to follow that the range of βούλευσις must coincide with that of the only remaining form of intellectual virtue, φρόνησις (1139f15-16). That this is so is clear from 1112b8-10 (τὸ βουλευέσθαι δὴ ἐν τοῖς ὡς ἐπὶ τὸ πολὺ, ἀδήλοισι δὲ πῶς ἀποβήσεται, καὶ ἐν οἷς ἀδιόριστον), and confirmed by 1140a24-b6 (a32: βουλευέται δ' οὐθεὶς περὶ τῶν ἀδυνάτων ἄλλως ἔχειν).

(v) Belief is qualified as either true or false, but προαίρεσις as right or wrong (h and k); προαίρεσις makes our characters good or bad- not moral belief (i). Likewise βούλευσις is that kind of deliberation employed in settling moral, practical problems: δοκεῖ δὴ φρονίμου εἶναι τὸ δύνασθαι καλῶς βουλευσάσθαι περὶ τὰ αὐτῷ ἀγαθὰ καὶ συμφέροντα (1140a25-26).

(vi) 'προαίρεσις concerns pursuit or avoidance, but belief, what things are, and how advantageous, and for whom' (j: 1112a4-5). Likewise οὐδὲ δὴ [εἶη βουλευτὸν] τὰ καθ' ἕκαστα, οἷον εἰ ἄρτος τοῦτο ἢ πέπεπται ὡς δεῖ· αἰσθήσεως γὰρ ταῦτα: 1113a1).

It is evident from this comparison that προαίρεσις and βούλευσις are indeed intimately connected. So intimately connected, in fact, that Aristotle actually feels it necessary to tell us how to distinguish them:

βουλευτὸν δὲ καὶ προαιρετὸν τὸ αὐτό, πλὴν ἀφωρισμένον ἤδη τὸ προαιρετόν· τὸ γὰρ ἐκ τῆς βουλῆς προκριθὲν προαιρετόν ἐστιν. (NE 1113a2-5)

Somewhat freely, I translate: 'What βούλευσις has settled on is the same as what προαίρεσις settles on, except that when προαίρεσις settles on something, it is already predeterminate. For what one's deliberating (βουλή) has already (προ-) adjudged, is what προαίρεσις settles on.' (Note, again, Aristotle's insistence on the use of the προ- prefix in προκριθὲν.) The same point is made in a different way by 1139a32:

πράξεως μὲν οὖν ἀρχὴ προαίρεσις (ὅθεν ἢ κίνησις ἀλλ' οὐχ οὐ
ἐνεκα), προαιρέσεως δὲ ὄρεξις καὶ λόγος ὁ ἐνεκά τινος.

The point of these passages is twofold. First: that the efficient cause of the πράξις- the 'action on προαίρεσις', as I have been calling it- is the προαίρεσις. And second: that there is something which precedes προαίρεσις as *its* efficient cause, just as προαίρεσις precedes action on προαίρεσις. 1113a2-5 implies that this is βούλευσις; 1139a32, that this is ὄρεξις καὶ λόγος ὁ ἐνεκά τινος, 'appetition plus reasoning directed to some end'.

This suggests that βούλευσις is *equivalent* to 'appetition plus reasoning directed to some end'. Bear in mind the NE III.2-3 doctrine that προαίρεσις and βούλευσις are of means, not ends, and compare 1112b13-14,16:

οὔτε γὰρ ἰατρὸς βουλευέται εἰ ὑγιάσει, οὔτε ῥήτωρ εἰ
πέσει... ἀλλὰ θέμενοι τέλος τι, πῶς καὶ διὰ τίνων ἔσται
σκοποῦσι...

(How do we translate θέμενοι here, given that (as I have argued) the establishment of first principles for reasoning is not necessarily itself an irrational process? Not as 'taking for granted', with the Loeb, but as 'positing' or 'laid down'. The doctor *qua* doctor does not deliberate about what end to seek to bring about; the τέλος of her actions is already settled for her by the very fact that she is a doctor. But a human being- *qua*

human being, as one might say- might well deliberate about *whether to be a doctor*; or indeed about whether to act, in some situation, *qua* doctor or (say) *qua* soldier. As we have seen, there are both analogies and disanalogies between a τέχνη and the ποίησις which is the ἔργον it must involve, and ἀρετή ὅλως, overall virtue, and the πράξις which it may or may not involve.)

Anyway, we may gloss 1139a32 as meaning that what happens in βούλευσις is that an end which the agent has already decided upon (by some other process which is not, strictly, βούλευσις, or at any rate not *this* βούλευσις) is converted into a προαίρεσις, a *decision to act*. This happens by way of the process which *is*, strictly, βούλευσις. By a mental searching, means to the end in question are found: πῶς καὶ διὰ τίνων ἔσται σκοποῦσι.

The best and most detailed description of this process is given at NE 1112b16-27. βούλευσις, says Aristotle, is a process of working from a first term, namely the τέλος which has been posited, to a final term, namely the beginning of the πράξις: τὸ ἔσχατον ἐν τῇ ἀναλύσει πρῶτον εἶναι ἐν τῇ γενέσει (b24).

(Note that, on this account, the βούλευσις *includes* the προαίρεσις, and is the efficient cause, not of the προαίρεσις as at 1139a32-33, but of the *action on* προαίρεσις, the πράξις itself. The cause of the trouble here may be the ambiguity of Aristotle's use of προαίρεσις, between 'decision to act' (so 1139a32) and 'action on a decision to act' (so, apparently, 1112a13-16).)

The deliberator, in 'examining how and by what means [the end] will be brought about' (b16), works from a first term, the τέλος, to a last term, the πρᾶξις or the beginning of the πράξις, by the quickest method possible- if any method is possible. If several methods are available, the deliberator asks herself 'Which is the easiest and best of these methods?':

διὰ πλειόνων μὲν φαινομένου γίνεσθαι διὰ τίνος ῥᾶστα καὶ κάλλιστα ἐπισκοποῦσι... (b17)

In the case where only one method is available- and, presumably, in the case where the deliberator has decided which is the best of several methods- the deliberator asks two further questions: 'How will this method bring about the end? How is this method itself to be brought into operation?':

...δι' ἑνὸς δ' ἐπιτελουμένου, πῶς διὰ τούτου ἔσται, κακεῖνο διὰ τίνος... (b18)

And so the deliberator works back to some performance or movement within his power which is immediately available for him to do without doing anything else first- and does it. But if at any stage in this process it becomes clear that in fact no method is available, the deliberator abandons the project:

...ἕως ἂν ἔλθωσιν ἐπὶ τὸ πρῶτον αἴτιον, ὃ ἐν τῇ εὐρέσει ἔσχατόν ἐστιν... ἐὰν δὲ δυνατόν φαίνεται, ἐγχειροῦσι πράττειν.
(b19, b27)

...κἂν μὲν ἀδυνάτῳ ἐντύχῳσιν, ἀφίστανται. (1112b25)

This performance or movement is the last term in the deliberation, and the first step (τὸ πρῶτον αἴτιον) in the performance of the action (b24). The use of the word αἴτιον here reminds us again that this account can be seen as explaining the 1139a32 remark, that the deliberation is the *efficient cause* of the action. The process of deliberation terminates in something which can be done as the beginning of a process of action: what Aristotle calls τὸ πρῶτον αἴτιον, and what David Charles (1984, Ch.2) calls a 'basic action'. The same point is suggested by Aristotle's formula (1139a5) about προαίρεσις, that it is ἡ ὀρεκτικὸς νοῦς ἢ ὄρεξις διανοητική. He repeats it in another form at 1113a5-6:

παύεται γὰρ ἕκαστος ζητῶν πῶς πράξει ὅταν εἰς αὐτὸν ἀνάγκη τὴν ἀρχήν, καὶ αὐτοῦ εἰς τὸ ἡγούμενον· τοῦτο γὰρ τὸ προαιρούμενον.

This analysis of Aristotle's account of deliberation suggests three points.

First: to give an account of deliberation, as I have just done, is itself to give an account of the combination of unhindered appetite (ὄρεξις for a τέλος) with relevant knowledge (awareness of the appropriate method for obtaining the τέλος, and of the best means to put that method into practice). It was, of course, this combination which I suggested was what

...κἄν μὲν ἄδυνάτῳ ἐντύχῳσιν, ἀφίστανται. (1112b25)

This performance or movement is the last term in the deliberation, and the first step (τὸ πρῶτον αἴτιον) in the performance of the action (b24). The use of the word here reminds us again that this account can be seen as a response to the 1139a32 remark, that the deliberation is the last term in the action. The process of deliberation is the last term in the action which can be done as the beginning of the action. Aristotle calls τὸ πρῶτον αἴτιον, and what we call (Ch.2) calls a 'basic action'. The same word is used in Aristotle's formula (1139a5) about πρῶτον αἴτιον. ὁρεκτικός νοῦς ἢ ὄρεξις διανοητική. He repeats this at 1113a5-6:

παύεται γὰρ ἕκαστος ζητῶν πῶς πράξει ὅταν εἰς αὐτὸν ἀνάγκη τὴν ἀρχήν, καὶ αὐτοῦ εἰς τὸ ἡγούμενον· τοῦτο γὰρ τὸ προαιρούμενον.

This analysis of Aristotle's account of deliberation suggests three points.

First: to give an account of deliberation, as I have just done, is itself to give an account of the combination of unhindered appetite (ὄρεξις for a τέλος) with relevant knowledge (awareness of the appropriate method for obtaining the τέλος, and of the best means to put that method into practice). It was, of course, this combination which I suggested was what

constituted the rationality of a voluntary action- whether the deliberation in question is actual, as in action on προαίρεσις, or merely potential, as in τὸ ἐκούσιον, the merely voluntary.

Second: we now have an account of something which may have seemed obscure thus far, namely what it means to say that some behaviour is performed '*because of*' a combination of unhindered appetite and relevant knowledge. 'The last step of the deliberation is the first step of the action' (1112b24): the πρῶτον αἴτιον is *both* the termination of the process of deliberation, *and* the commencement of the process of action- and hence is itself the causal link between thought and action. (Cp. §4d below, on the nature of the conclusion of a 'practical syllogism'). Thus to say that some behaviour B is performed '*because of*' the combination of unhindered appetite and relevant knowledge in deliberation is to say no more and no less than that B satisfies the rationality condition of voluntary action.

And third: if this is what deliberation is, then it would seem that a more precise account of the nature of deliberative reasoning might be obtained by closer examination of the notion of the 'practical syllogism'. To this I now turn.

4. Practical Reasoning and the 'Practical Syllogism'

The suggestion is that an account of practical reasoning in

the form of the 'practical syllogism' is at the heart of Aristotle's account of deliberation, which itself is central to Aristotle's account of action on προαίρεσις, the most fully explicit form of voluntary action. What, then, is the Aristotelian 'practical syllogism'? In particular, how far is it *inference* in the strict sense? Answering this question will involve me in a brief examination (§4a) of standard Aristotelian syllogisms, so as to compare them (§4b) with practical syllogisms.

First, a note on the *name* 'practical syllogism'.

'The expression ['practical syllogism'] is not a translation or transliteration of any expression Aristotle ever uses... The Greek words συλλογισμοὶ τῶν πρακτῶν [NE 1144a32] are taken as the Aristotelian authority for the use of the expression... But to take [these words] as a unit of expression is to misread the Greek.'

(Kenny 1979, p.111)

NE 1144a31-33 reads: οἱ γὰρ συλλογισμοὶ τῶν πρακτῶν ἀρχὴν ἔχοντες εἰσιν, ἐπειδὴ τοιόνδε τὸ τέλος καὶ τὸ ἄριστον... . As Kenny says, ἔχοντες εἰσιν cannot possibly be the Greek for 'are such as to have'. Hence rightly dividing the vital words here gives us this:

οἱ γὰρ συλλογισμοὶ | τῶν πρακτῶν ἀρχὴν ἔχοντες | εἰσιν...

and not, as Ross and Rackham apparently held, this:

οἱ γὰρ ἰ συλλογισμοὶ τῶν πρακτῶν ἰ ἀρχὴν ἔχοντες εἰσιν ἰ... ,

and it follows that we may render this phrase as:

'The syllogisms which contain starting points of actions are like this...'

We may not, however, render it (with Ross 1925) as:

'Practical syllogisms involve starting points of this sort...'

nor yet (with Rackham 1926) as

'For deductive inferences about matters of conduct always have a major premiss of the form...'

Which noted, I shall nonetheless follow the usual practice in using (*faute de mieux*) the phrase 'practical syllogism'-henceforth without scare quotes- to describe the particular 'sums' or pieces of inference which Aristotle, clearly enough, believes to be the basic components of the phenomenon of practical reasoning in general.

4a. The Standard Aristotelian Syllogism

To understand Aristotle's account of the practical syllogism, we need to compare it very briefly with his account of the standard, 'theoretical' or 'abstract', syllogism. Aristotle's account of the standard syllogism is part of his account of general abstract reasoning, which in turn is part of his account of argumentation: i.e., his account of how to give persuasive arguments (πίστεις, dAR 1354a16) or demonstrations (ἀπόδειξεις, APr 24a11). At dAR 1354a1 ff., he distinguishes two types of argument:

(i) *Logical* arguments (which can be either 'demonstrations' or 'dialectical arguments'), which depend on appealing solely to reason.

(ii) *Rhetorical* arguments. These depend on appealing to 'the moral character of the speaker, [by] some way of making the listener feel, and [by] the speech (ὁ λόγος) itself, through its proving or seeming to prove its thesis' (dAR 1356a1-4).

Aristotle divides logical argumentation into demonstration and dialectic (APr 24a23 ff.), or, perhaps equivalently, into syllogism and induction (APr 68b13-14; NE 1139b26: 'All teaching is *either* through induction *or* through syllogism').

What is Aristotle's contrast between induction and

syllogism? Whether induction is characterised as narrowly and tightly as it seems to be at APr 68b15-29 (and NE 1098b4), or more loosely so that it includes definition, habituation and the activities of αἴσθησις, φαντασία, or νοῦς (dMA 701a32-33), Aristotle makes it clear that it is induction's job to supply the first principles of syllogism:

τῶν ἀρχῶν δ' αἰ μὲν ἐπαγωγῇ θεωροῦνται, αἰ δ' αἰσθήσει, αἰ δ' ἐθισμῷ τινί, καὶ ἄλλαι δ' ἄλλως. (NE 1098b4)

Now Aristotle does not believe that induction (in the strict or in the loose sense) is rationally justifiable in the argumentative form of the syllogism, to which induction is prior and by which induction is presupposed. But this does not commit him to saying that induction is not, in any sense, rationally justifiable. Consider the sentence just before the last quotation, NE 1098b1-3:

οὐκ ἀπαιτητέον δ' οὐδὲ τὴν αἰτίαν ἐν ἅσασιν ὁμοίως, ἀλλ' ἱκανὸν ἔν τισι τὸ ὅτι δειχθῆναι καλῶς, οἷον καὶ περὶ τὰς ἀρχάς· τό δ' ὅτι πρῶτον καὶ ἀρχή.

For Aristotle different kinds of ἀρχή are justified in different ways. Does this passage go against that view? Not if we translate it, as we quite legitimately may, 'Nor again must we in all matters demand an explanation *of the same sort* (ὁμοίως)'. (Again, on this view, τό δ' ὅτι πρῶτον καὶ ἀρχή need not be, with Rackham (1926), 'The fact is *the* primary thing and *the*

principle'; it can equally well be 'The fact is a primary thing and a principle'. It is not as if there were only one kind of ἀρχή.)

The syllogism is the paradigm of rational justification, and yet it is not self supporting. For the formation of its principles, and hence for its very existence, the syllogism depends crucially on other forms of rational justification— which, be it noted, are both *other* forms, and forms of *rational justification*.

The syllogism is the τελείος form of theoretical reasoning: i.e., what all reasoning 'aims at being' (or creating) is the syllogistic form (APr 24b23-25). Only syllogisms are or can be completely and rigorously valid. So induction lacks the formal cogency of syllogism. Yet what induction lacks in formal cogency, it makes up (says Aristotle) in immediacy and subjective certainty: 'If [a reasoner] is not more certain of his first principles than of his conclusion, he will have his scientific knowledge accidentally' (NE 1139b34-35). 'By nature, the syllogism by the middle term [*sc.* the standard syllogism] is prior and more knowable (γνωριμώτερος, which also means 'nobler'). But to us, the [syllogism] by induction is more obvious (εναργέστερος)' (APr 68b36-37).

Aristotle saw that induction too plays a crucial role in human reasoning, a role which syllogistic inference could not possibly fill. That role is the establishment, whether by

perception, definition, or induction strictly so called, of the first principles of those syllogisms. This is a necessity *prior* to the establishment of any syllogisms at all: ἡ μὲν δὲ ἐπαγωγή ἀρχῆς ἐστὶ καὶ τοῦ καθόλου, ὁ δὲ συλλογισμὸς ἐκ τῶν καθόλου (NE 1139b28-30). Admittedly some syllogisms do indeed, as the 'rationalists' were claiming, derive their first principles from other syllogisms (call them *derivative syllogisms*). But at least some syllogisms (*prime syllogisms*) must, Aristotle thinks, be dependent on first principles arrived at not by syllogism but by induction. Eventually, in any reasoning, 'there will be a stop' (NE 1142a30)- the syllogising will come to rest on a first principle known by induction.

The doctrine of the *Analytics* is that we can classify all *fully* rigorous logical arguments as what Aristotle calls 'full' or 'perfect' syllogisms: ([ὁ] τέλειος συλλογισμὸς). Less rigorous arguments, but which are still clearly logical or pseudo-logical in character, will be 'imperfect' (ἀτελής) syllogisms- *degenerate* types of syllogism. (Syllogisms of proper form can be less than fully cogent too, of course. They may involve falsehoods and improper transitions.) Other less than τελείος forms of argumentation are the two rhetorical forms, enthymeme (APr 70a11-23) and παράδειγμα (APr 68b38-69a13): from a logical point of view, these turn out to be imperfectly formed syllogism and induction respectively.

Full syllogisms are either simple syllogisms; or they are what are called *sorites*, involving a series of transitions from

one simple syllogism to another, in which one premiss of the second syllogism is the conclusion of the first, of the third, the conclusion of the second- and so on. Aristotle, I take it, believes that all sorites are resolvable into simple syllogisms (APr 42a32-b26 seems to suggest this, without actually saying it).

There is on Aristotle's view a definite number of argumentative forms, some of which will always be valid and some of which will always be invalid (provided, of course, that we are dealing with *well-formed* examples of these forms). Aristotle's project, in the *Analytics*, is to classify these argumentative forms. If he succeeds, anyone who wants to show that a given dialectical argument is rigorously valid will need only to establish (i) that it is not a piece of induction, (ii) that it is an example of one of the rigorously valid or invalid forms of syllogism, and (iii), if (ii) is not sufficient to show its validity, that it is a well-formed example of a rigorously valid form of syllogism. Aristotle wants to give us a 'look-up table' of all the kinds of good argument there are- and some account of why only these kinds of argument are good.

4b. Comparison with the Practical Syllogism

How far, then, does this general account of the syllogism apply to the practical syllogism? To quote Anthony Kenny again:

'Practical syllogisms are not, and most of them do not even look like, syllogisms. A syllogism should consist of two premisses and a conclusion, all three grammatically of subject-predicate form; it should contain three terms, one of which ("the minor") occurs in subject place in the conclusion, one of which ("the major") occurs in predicate place in the conclusion, and the third of which ("the middle") occurs in both premisses but not in the conclusion... A traditional syllogism of this kind is something very unlike the patterns of non-theoretical reasoning in Aristotle... We find practical inferences involving two, three and four premisses; the conclusion is never a straightforward subject-predicate sentence; the premisses are often of conditional form so that the whole looks more like an exercise in propositional calculus than in syllogistic... one must realise that a "practical syllogism" is something even in appearance very different from a syllogism in Barbara or an inference of the form "All Xs are Ys; A is an X; therefore A is a Y."

(Kenny 1979, p.112)

It is certainly true that Aristotle's remarks on and examples of practical syllogisms, when compared with his remarks on and examples of abstract syllogisms, seem puzzling and unhelpful. The sustained and intricate account of the abstract syllogism in the logical works stands in stark contrast to Aristotle's scattered and informal remarks on the practical syllogism, and his desultory and apparently carelessly-chosen examples. It does not

look easy to extract a single canonical form from these.

Nonetheless, although the practical syllogism is undeniably unlike other Aristotelian syllogisms in some respects, it is not *totally* unlike. There is room for doubt that the practical 'syllogism' is truly a syllogism, when we consider Aristotle's definition of 'syllogism' at APr 24b19-20: συλλογισμός δέ ἐστι λόγος ἐν ᾧ τεθέντων τινῶν ἕτερον τι τῶν κειμένων ἐξ ἀνάγκης συμβαίνει τῷ ταῦτα εἶναι. But on the other hand, there is no doubt at all that Aristotle thinks that the practical syllogism expresses *some* kind of logic.

My line will be that Aristotle viewed the relation between standard syllogisms and practical syllogisms as no more, and no less, than an analogical relation. That the relation between them is not simply identity of kind, but not simply dissimilarity either, seems to be expressed by Aristotle's language about practical syllogisms. If the relation was identity of kind, we might expect Aristotle's commonest term for 'practical syllogism' to be the heavily theory-charged term συλλογισμός. But, in fact, he uses συλλογισμός of practical reasoning on only two occasions in the NE (1142b23, 1144a31; but cp. also 1149a33 ὡςπερ συλλογισάμενος). Again, if the relation was anything approaching simple dissimilarity, we might expect Aristotle to prefer the very neutral, non-committal term λόγος (which he uses of

practical reasoning in six places: 1112a16, 1149a26, 1149b1-3, 1150b28, 1151a29). But neither of these terms is Aristotle's commonest term for 'practical syllogism'; this is λογισμός (NE 1111a34, 1117a21, 1119b10, 1141b14, 1142b19, 1145b11-12, 1146a33, 1149b35, 1150b24). Moreover, unlike λόγος, λογισμός never seems to have any other sense in the NE. (The only possible exception is 1141b14, but even this seems dubious.)

This terminological survey has a moral: the reasoning of practical syllogisms is not so loose as to deserve no more than the prosaic name λόγος; but not so rigorous as to deserve the special name συλλογισμός. It is somewhere in between. Adding the scientific looking suffix -ισμος promotes λόγος; dropping the intensive prefix συν- demotes συλλογισμός. By either route, or both, one arrives at λογισμός. The word was not of course an Aristotelian coinage; but its shape was very convenient for expressing what he wanted.

I will now substantiate my suggestion that the relation between the standard syllogism and the practical syllogism is an analogical one. The right place to start this account is with the examples of practical syllogisms which Aristotle actually gives.

So here, in my own very literal English, are all twelve (as I count them) of Aristotle's formal or specific examples of complete or readily completable practical syllogisms, stripped as far as possible of the commentaries in which they are embedded.

(1) '[Belief and reasoning about the universal] says that such a person should do such a sort of thing; [belief and reasoning about the particular fact says] that this thing now is a thing of that sort, and that I myself am such a person.' (dA 434a17-20)

(2) '...As when he thinks that every man should walk, and that he is a man, he at once walks...' (dMA 701a14)

(3) '...or if he thinks that no man should walk now, and he is a man, he at once is still.' (dMA 701a15)

(4) "'I should make something good"; "A house is something good"; he makes a house at once.' (dMA 701a16-17)

(5) "'I need a covering"; "A cloak is a covering"; "I need a cloak". "I should make what I need"; "I need a cloak"; "I should make a cloak".' (dMA 701a17-20)

(6) "'I should drink", says desire. "This is a drink", says perception or imagination or νοῦς. At once he drinks.' (dMA 701a32-33)

(7) "'Since this is health, it is necessary, if the patient is to be healthy, that so and so should be the case, such as 'homogeneity'. If there is to be 'homogeneity', there must be

heat..." -and so he always continues thinking until he arrives at the last term, what he himself can do.' (Mph 1032b7-9)

(8) '- "If he is to be healthy, he must be made homogeneous." - "So what counts as homogeneity?" - "This; and this will come about if he is heated." - "What counts as being heated?" - "This; and this is present in potential [i.e., is possible] and is already up to him."' (Mph 1032b18-22)

(9) '[One can be wrong] either that all "heavy water" is unwholesome, or that this is "heavy water".' (NE 1142a23)

(10) '...He may know that dry foods are good for every man, and that he himself is a human, or that Food X is a dry food; but that this is Food X, he may either not know or not actualise.' (NE 1147a5-7)

(11) 'If every sweet thing should be tasted, and this thing is sweet- being some particular item- it is necessary for a person, who is able and is not prevented, at once also to taste it...'. (NE 1147a29-31)

(12) '...So when the universal premiss is present which forbids tasting [whatever is sweet], but also the [universal premiss] that "every sweet thing is pleasant", and [the particular premiss] that "this thing is sweet", and this latter is operative when desire happens to be present; then the [first universal premiss] says that one should flee this, but the desire leads one on.' (NE 1147a31-34)

What are we to make of these examples of practical syllogisms, when we compare them with standard syllogisms? There are five main problem areas:

- (a) Do practical syllogisms have only two premisses, one major and one minor?
- (b) What is the 'major premiss' of a practical syllogism?
- (c) What is the 'minor premiss' of a practical syllogism?
- (d) What is the conclusion of a practical syllogism?
- (e) Are practical syllogisms, in fact, *valid*?

(a) This question divides into two sub-questions: (a1) How many premisses does a practical syllogism have? (a2) Of what sorts?

On (a1), DMA 701a12 (ἐκ τῶν δύο προτάσεων...) is clear enough; but DMA 701a23-25 (αἱ δὲ προτάσεις αἱ πολιτικαὶ διὰ δύο εἰδῶν γίνονται) clouds the picture by not telling us for sure that only *one* premiss of each εἰδὸς comes to be in a practical syllogism. My examples (1-4), (6), (9), and (11) support the DMA 701a12 view; but (5), (7), (8), (10) and (12) go against it. And NE 1147a4-10 argues explicitly that there are two (or is it three?) major premisses in each practical syllogism:

διαφέρει δὲ καὶ τὸ καθόλου· τὸ μὲν γὰρ ἐφ' ἑαυτοῦ τὸ δ' ἐπὶ τοῦ πράγματος ἐστίν, οἷον ὅτι παντὶ ἀνθρώπῳ συμφέρει τὰ ξηρά, καὶ ὅτι αὐτὸς ἄνθρωπος, ἢ ὅτι ξηρὸν τὸ τοιόνδε... (1147a4-7)

Consider, however, the difference between standard 'syllogisms' which are simple syllogisms and those which are actually sorites. It is, I suggested, Aristotle's conviction about standard syllogisms that all sorites can be analysed as

compounded out of simple syllogisms. Why not apply this idea to practical syllogisms also? Suppose we had formalised example (5) as follows:

- (5') 1. (X thinks:) X needs a covering.
2. (X thinks:) A cloak is a covering.
3. (X thinks:) X needs a cloak.
4. (X thinks:) What X needs, X must make.
5. (X thinks:) X must make a cloak.

Given the point about sorites, we might renumber (5'5) as (5'6) and add a new (5'5), of identical content to (5'3):

- (5') 5. (X thinks:) X needs a cloak.

(5'1-3) and (5'4-6) can now be seen as the two simple syllogisms out of which (5') is compounded as a sorites. For clarity, we may write the sorites out, showing the transferred conclusion, and with its constituent simple syllogisms marked off, like this:

- (5') 1. (X thinks:) X needs a covering.
2. (X thinks:) A cloak is a covering.
3. (X thinks:) X needs a cloak.
4. (X thinks:) What X needs, X must make.
5. (X thinks:) X needs a cloak.
6. (X thinks:) X must make a cloak.

A similar treatment can be applied to the other cases which show just the same problem, (7), (8) and (10); with, however, the added complication that these are three-stage, not two-stage, sorites. Given that (7) and (8) are essentially the same syllogism, we might formalise that syllogism thus:

- (7') 1. (X thinks:) The patient Y is to be healthy.
2. (X thinks:) Health is 'homogeneity' (ὁμολότης).
3. (X thinks:) X must increase Y's 'homogeneity'.
4. (X thinks:) X must increase Y's 'homogeneity'.
5. (X thinks:) Y's 'homogeneity' will be increased if Y is heated.
6. (X thinks:) X must heat Y.
7. (X thinks:) X must heat Y.
8. (X thinks:) If Y is massaged, Y will be heated.
9. (X thinks:) X must massage Y.

Example (10), likewise, will give us this sorites:

- (10') 1. (X thinks:) To be healthy, X must eat what is digestible and wholesome.
2. (X thinks:) Light meat is digestible and wholesome.
3. (X thinks:) X should eat light meat.
4. (X thinks:) X should eat light meat.
5. (X thinks:) Chicken is light meat.

6. (X thinks:) X should eat chicken.

7. (X thinks:) X should eat chicken.

8. (X thinks:) This is chicken.

9. (X thinks:) X should eat this.

(a2) What sorts of premisses does a practical syllogism have?

It appears from the *Analytics* that standard syllogisms have one minor premiss or premiss 'of the particulars' (τὰ καθ' ἕκαστα), and one major premiss or premiss 'of the universal' (τὸ καθόλου). In the NE, Aristotle does use the terms τὸ καθόλου and τὰ καθ' ἕκαστα when referring to the premisses of practical syllogisms: so 1143b3-4, 1147a1-10, 27 ff., and cf. 1135a6-8: τῶν δὲ δικαίων καὶ νομίμων ἕκαστον ὡς τὰ καθόλου πρὸς τὰ καθ' ἕκαστα ἔχει· τὰ μὲν γὰρ πραττόμενα πολλά, ἐκείνων δ' ἕκαστον ἓν. The same terms are employed in the dA (434a17-18). But in the dMA, Aristotle uses a different pair of terms:

αἱ δὲ προτάσεις αἱ ποιητικαὶ διὰ δύο εἰδῶν γίνονται, διὰ τε τοῦ ἀγαθοῦ καὶ τοῦ δυνατοῦ... (701a23-25)

A division of premisses into 'premisses of the good' and 'premisses of the possible' seems to suggest a rather different

picture. Which

(b) brings me to an examination of the practical syllogism's major, or universal, premiss, or premiss of the good- whichever we should call it. The term 'major premiss' should, perhaps, be reserved for standard syllogisms, so that practical syllogisms have neither major nor minor premisses. But considering some of the universal premisses in the examples above, neither of the other two pairs of names seems to fit either: 'Every man should walk' (2), 'I need a covering' (5), '"I should drink", says desire' (6), 'Dry food is good for every man' (10). In what sense are any of these premisses *of the good*? In what sense are (5) and (6) *universal* premisses? (Until these questions are answered, I will call them first premisses.) Two important passages in particular seem hard to square with the idea that such propositions as these could be the first premisses of practical syllogisms, NE 1144a31-33 and dA 434a17-20.

(i) As we have seen, NE 1144a31-33 reads: οἱ γὰρ συλλογισμοὶ τῶν πρακτῶν ἀρχὴν ἔχοντες εἰσιν, ἐπειδὴ τοιόνδε τὸ τέλος καὶ τὸ ὄριστον... . The Loeb here has: 'For deductive inferences about matters of conduct always have a major premiss of the form 'Since the End or Supreme Good is so and so...'' (Loeb, p.368). This is a miracle of mistranslation, even if we ignore the συλλογισμοὶ τῶν πρακτῶν business: note the completely baseless introduction of

'always', and the assumption that τὸ τέλος καὶ τὸ ἄριστον can and should be rendered in the Kantian upper-case. Nonetheless, even if the Greek does not make anything like such a strong claim as the Loeb- as it plainly does not- this passage seems at odds with the idea that anything but a very basic moral first principle could be an ἀρχή of a practical syllogism.

(ii) dA 434a17-20: ἡ μὲν καθόλου ὑπόληψις καὶ λόγος, ἡ δὲ τοῦ καθ' ἕκαστα (ἡ μὲν γὰρ λέγει ὅτι δεῖ τὸν τοιοῦτον τὸ τοιόνδε πράττειν, ἡ δὲ ὅτι τόδε τὸ νῦν τοιόνδε, κάγω δὲ τοιόσδε)...

How is one to reconcile this with NE 1144a31-33 and the examples above? The first premisses of (2, 3) ('Every/ no man should walk'), (7, 8) 'The patient is to be healthy', (9) ('All "heavy water" is unwholesome'), (10) ('Dry food is good for every man'), and (11) ('Every sweet thing should be tasted') may fit the pattern laid down by dA 434a17-20, but are, it seems, too narrow to be examples of general statements that ἐπειδὴ τοιόνδε τὸ τέλος καὶ τὸ ἄριστον..., as NE 1144a31-33 apparently requires them to be. On the other hand, the first premiss of (4) ('I should make something good'), might fit the NE formula, but is surely too broad to be a statement ὅτι δεῖ τὸν τοιοῦτον τὸ τοιόνδε πράττειν, as the dA requires. As for the first premisses of (5) ('I need a covering') and (6) ('"I should drink", says desire'), these do not seem to fit with either the dA or the NE doctrine.

A reconciliation of this conflicting evidence can, however,

be brought about. Perhaps unsurprisingly, it turns on a correct translation and interpretation of 1144a31-33. The correct translation is something like this:

'The syllogisms which contain the starting point of actions run: "Since such and such is the goal and the best..."'

The correct interpretation depends on these two conditions. First, that 'the goal and the best' should not be given too exalted an understanding (and certainly nothing which deserves capitalisation); for Aristotle carries on: ...ὅτι δὴ ποτε ὄν· ἔστι γὰρ λόγου χάριν τὸ τυχόν (33-34). Second, that it be understood that the kind of moral ἀρχή with which Aristotle is concerned is, as correct translation shows, not some moral "ἀρχή ἁπλως", but the τῶν πρακτῶν ἀρχή.

1144a31-33 does *not*, as the Loeb suggests, mean that a practical syllogism must start with a moral ἀρχή *tout court*, a completely general statement about what is good. Rather, one need only start with a τῶν πρακτῶν ἀρχή, a statement of the good which is to be sought in some particular case of action. Hence, rightly interpreted, 1144a31-33 is not out of line with dA 434a17-20, but in complete accord with it: for it too teaches that '[Belief and reasoning about the universal] says that such a person should do such a sort of thing'.

This account leaves out the first premisses of (4) ('I should make something good'), (5) ('I need a covering') and (6) ('I

should drink", says desire'). 'I should make something good' is, I think, simply an aberration; this is nearer a moral ἀρχή *tout court*, a completely general statement about what is good, than a τῶν πρακτῶν ἀρχή. It is too vague to do the practical work which Aristotle requires of a τῶν πρακτῶν ἀρχή. (In fact it is very difficult to see how this thought could be the starting point of the chain of reasoning which follows it in (4). At the very least, we need to amend it to: 'In this situation I should make some appropriate good'.)

'I need a covering' and 'I should drink' look more like particular premisses than first premisses. Perhaps Aristotle wants us to assume sorites in which these are intermediate conclusions, as for example:

(i) When it rains, all men should avoid getting wet

(ii) It is raining here and now

(iii) All men here and now should avoid getting wet

(iv) All men here and now should avoid getting wet

(v) I am a man here and now

(vi) I should avoid getting wet here and now

(vi) I should avoid getting wet here and now

(viii) If I have a covering, I will avoid getting wet here and now

(ix) I need a covering

-(ix) emerges from this sorites, and one can see that 'I should drink' might emerge from something similar. But it would be a mistake to call (ix) a first premiss without careful qualification. It would be a mistake to claim that (ix) and its ilk are, strictly, τῶν πρακτῶν ἀρχαί (as I use the phrase), statements of the good which is to be sought in some particular case of action. These are not statements of the *good* as such at all. But it is clear how such statements might be brought back to depend on statements of the good at some earlier stage in what will be a rather complex sorites.

The upshot of this analysis, then, is that the first premiss of a practical syllogism is an ἀρχή τῶν πρακτῶν, that is, a statement of the good which is to be sought in some particular case of action. It is not, in spite of the appearance of 1144a31-33, a moral ἀρχή *tout court*, a completely general statement of what is good.

This conclusion apparently suggests that moral ἀρχαί *tout court* lurk behind ἀρχαί τῶν πρακτῶν, as their ultimate rational vindication; and (as at 1135a6-9) Aristotle often seems to allow this kind of talk. Indeed, if the practical syllogism, just like the standard syllogism, rests on first principles attained by 'induction', it apparently ought to follow that we can specify the content of those moral first principles, those ἀρχαί *tout court*, just as we can specify the content of (e.g.) mathematical first principles (1142a27 ff.). However, it is an important point about Aristotle's moral outlook that, if there are any such moral

first principles at all, he seems very coy about giving us examples- as noted in Ch.2. Without the presence in a person of *good character* (1139a33-34), it seems that no moral first principles in the strict sense will be available to that person; except possibly those unhelpful truisms of which the first premiss of (4) would be such a nice example if it read, as it so nearly does, 'All people should make *whatever* is good'. (For the suggestion that the (ultimate) moral principles are κατὰ τὰς ἠθικὰς ἀρετὰς, in accordance with (or even identical with?) the virtues, cp. NE 1178a18.)

(c) Given that this is the nature of the first premiss of an Aristotelian practical syllogism, and given what was established in (a) about the number and kind of the practical syllogism's premisses, it should now be plain what the second premiss, the 'premiss of the possible', will typically be like. I have cleared the way for the view of dA 434a17-20 concerning the first premiss of the practical syllogism. But there is far less conflict about the nature of the second premiss: it is not as hard to reconcile (for example) the claims that the second premiss is a [δόξα] περὶ τῶν καθ' ἕκαστα, ὧν αἰσθησις ἤδη κυρία (NE 1147a26) and that it is the premiss which says ὅτι τόδε τὸ νῦν τοιόνδε, κἀγὼ δὲ τοιόσδε (dA 434a20).

The second premiss of the practical syllogism is concerned with the *application* of the moral belief or other pro-attitude expressed by the first premiss. It is the second premiss which does the vital work in making reason practical, in getting us

from a moral belief (the first premiss) to an actual deed (the conclusion). Aristotle's stress on the means, not the ends, side of practical reasoning in NE III.2-3, and his remarks that it is ἡ τελευταία πρότασις which is κυρία τῶν πράξεων (1147b10), and that διάνοια δ' αὐτῇ οὐθὲν κινεῖ, ἀλλ' ἡ ἕνεκά του καὶ πρακτικὴ (1139a35), are all different ways of pointing to the crucial importance of the second premiss of the practical syllogism.

On the basis of dA 434a17-20, then, it seems that the paradigm of the practical syllogism will be a piece of reasoning of this form:

(i) (ἡ μὲν καθόλου ὑπόληψις καὶ λόγος:) δεῖ τὸν τοιοῦτον τὸ τοιόνδε πράττειν.

(ii) (ἡ δὲ τοῦ καθ' ἕκαστα [ὑπόληψις καὶ λόγος]:) τόδε τὸ νῦν τοιόνδε, κἀγὼ δὲ τοιόσδε.

(iii) (τὸ συμπέρασμα, 1139b34:) [ἐμέ δεῖ τόδε τὸ νῦν πράττειν.]

Notice three things about this paradigm of practical reasoning. First, how readily the allegedly simple syllogism splits into a two-stage sorites:

(i) (ἡ μὲν καθόλου ὑπόληψις καὶ λόγος:) δεῖ τὸν τοιοῦτον τὸ τοιόνδε πράττειν.

(ii) (ἡ δὲ τοῦ καθ' ἕκαστα [ὑπόληψις καὶ λόγος]:) τόδε τὸ νῦν τοιόνδε.

(iii) (τὸ πρῶτον συμπέρασμα:) δεῖ τὸν τοιοῦτον τόδε τὸ νῦν πράττειν.

(iv) (ἡ μὲν καθόλου ὑπόληψις καὶ λόγος:) δεῖ τὸν τοιοῦτον τόδε τὸ νῦν πράττειν.

(v) (ἡ δὲ τοῦ καθ' ἕκαστα [ὑπόληψις καὶ λόγος]:) κάγω δὲ τοιόσδε.

(vi) (τὸ δεύτερον συμπέρασμα:) ἐμὲ δεῖ τόδε τὸ νῦν πράττειν.

Second, notice that the desired distinction between first and second premisses- that the first should state a moral belief, or at least a pro-attitude, the second identify a case to which that attitude applies- is preserved in this sorites (compare (i) with (ii), and (iv) with (v)). The same distinction should be preserved in all practical syllogisms and sorites of canonical form- even at the stage of the sorites where the next first premiss turns out to be as specific and as infra-moral as 'I need a cloak'.

And third, notice how closely this kind of syllogism fits in with the view of Aristotelian ethics as based on the 'man of excellence', and his expertise in definition and in the application of definitions, which I argued for in the last chapter. First premisses like (i) and (iv) are indeed, in a perfectly straightforward sense, definitions; and second premisses like (ii) and (v) are indeed applications of these definitions. These first premisses are, in a way, universal, but they do not have the legalistic form of statements like 'All murder is wrong' or 'Always practise charity'. What they say is not so sharply defined, yet no less practically realisable: 'Such a sort of person should do such a sort of thing'. And in this context, I suggest, saying that τοιόσδε should act as prescribed

is virtually equivalent to saying that a 'man of excellence' should so act.

(d) What about the form of the conclusion of the practical syllogism? It is a much-debated question whether the conclusion of a practical syllogism is an action or a proposition. Charles (1984), for example, argues that it cannot be an action; the conclusion of a practical syllogism is a proposition, and only the presence of desire in addition to the presence of that conclusion can render the conclusion operative (Charles 1984, Ch.4). Anscombe famously argued that the conclusion of the practical syllogism was an *intention*, not an action (Anscombe 1957); and other writers have followed this sort of line. Other philosophers again (e.g. von Wright 1978) have argued that the conclusion of a practical syllogism is not a proposition, but an action. But the evidence which Aristotle gives us does not obviously favour either of these views over the other. Indeed, if one accepts that one must take either the one view or the other, Aristotle seems to contradict himself.

In the DMA, he apparently argues for the view that the conclusion of a practical syllogism is not a proposition, but an action: 'From the two premisses, the conclusion that results is the action. Thus, if someone thinks that "Every man ought to march", and that "He himself is a man", he marches *at once* (εὐθέως)' (DMA 701a14-15). In the NE, moreover, he adds that 'When one [opinion] results from these, it is necessary for the soul to *affirm* the consequent there [sc. in the standard

sylllogism], and [likewise necessary here], in practical [deliberations], *immediately to perform the consequent*' (NE 1147a27-29).

On the other hand, there is the doctrine of NE 1139a32-33, that *ᾄρεξις καὶ λόγος ὁ ἕνεκά τινος* is the cause of the *προαίρεσις*, and *προαίρεσις* in turn of the *πρᾄξις*. I have argued that *ᾄρεξις καὶ λόγος ὁ ἕνεκά τινος* is equivalent to deliberation, and that the practical syllogism is the form of deliberation. Hence what the practical syllogism causes (or at any rate, causes directly)- where 'what it causes' means 'what it terminates in', 'what its conclusion is'- will not be the *πρᾄξις* but the *προαίρεσις*. In support of that view, there is also the evidence of DMA 701a15-16 ('In both cases the agent does these things *unless something prevents him or compels him*') and the second δόξα of NE VII: καὶ ὁ αὐτὸς ἐγκρατῆς καὶ ἐμμενετικὸς τῷ λογισμῷ, καὶ ἀκρατῆς καὶ ἐκστατικὸς τοῦ λογισμοῦ (1145b11-12; cp. 1151a29-30). Aristotle endorses, cautiously at least, the view that *akrasia* involves or may involve a rejection of (the result of) a *λογισμός*. This clearly cannot mean the rejection of a *πρᾄξις*, since no *πρᾄξις* has yet been performed; so it must presumably mean the rejection of a *προαίρεσις*.

The evidence, then, conflicts. I suspect there are good reasons for this. As we have already seen (§3), Aristotle equivocates between (a) saying that the *βούλευσις* *includes* the *προαίρεσις*, and is the efficient cause of the *action on προαίρεσις*, the *πρᾄξις* itself (so 1112a13-16), and (b) saying

that the βούλευσις is separate from the προαίρεσις, and is the efficient cause of the προαίρεσις (so 1139a32-33). Likewise, therefore, he also equivocates between saying that the practical syllogism (which is the βούλευσις) leads to the προαίρεσις, which in turn leads to the action, and saying that the practical syllogism itself leads directly to the action.

Aristotle evidently wants to sit on this particular fence. He does not want to be forced into saying either that the conclusion of a practical syllogism is a proposition and therefore cannot be an action, or that it is an action and therefore cannot be a proposition. What he wants to establish is that the conclusion of the practical syllogism is a *hinge*, a causal link, between thought and action. So NE 1112a24-25 ('The last step in the analysis [made by reason] is the first step in the genesis [of action]') and Mph 1032b7-10: 'And thus he will always continue thinking (νοεῖ) until he arrives at what he himself can do- the last term'. But attempts to classify the conclusion of a practical syllogism as either simply a proposition, or simply an action, beg the question if they assume in advance that there can be no such link, that the movement from thought to action cannot be smooth and continuous.

For Aristotle, the conclusion of the practical syllogism is, or can be expressed in, both a proposition (such as an intention can give rise to) and an action. And why not? Generally speaking, it surely makes sense to suggest that an action can embody, incarnate, a proposition, just as much as a proposition can give

an interpretation of, a 'read-out' from, an action. (Think of mime artists, for instance, or the action that 'says' as clearly as any verbal remark 'I am disgusted by the taste of this water'.)

The conclusion of the practical syllogism will appear most obviously as an action in cases of $\eta\rho\acute{\alpha}\xi\iota\varsigma$ where there is little need to run through one's reasons for acting in a specific way—that is, where the deliberation involved in the action remains implicit, and the rationality of the action is potential rather than actual (52). It will appear most obviously as a proposition rather than an action in cases where the expression of the action form of the conclusion is thwarted, e.g. by compulsion or by *akrasia*, so that only the propositional form remains:

'It is normally only in the case where decisions are not acted on (in the case, for instance, of the incontinent) or where action is postponed that there is any point in distinguishing between the act and the decision.'

(Kenny 1979, p.143)

But there is, of course, nothing impossible about cases where the conclusion of the practical syllogism is expressed *simultaneously* as both action and proposition ('You need help,' I say to the tramp, giving him a fifty pound note). Certainly Aristotle sits on the fence on this question, and betrays some confusion and incoherence about how to answer it. But, confusions

and incoherences aside (and they are not irremediable), perhaps he was right to want to sit on it, if he is to succeed in presenting an account of 'how it is that a thinking agent sometimes acts and sometimes does not' (DMA 701a7).

(e) Finally, then, to the validity of the practical syllogism. Much has been written about this. One notable contribution is Hare (1969). Hare distinguishes between the necessary and sufficient logical or causal conditions of satisfying an imperative, meaning a universal imperative (which is what he thinks any moral statement amounts to). He applies this distinction to such imperatives as (A1) 'All men must march' (DMA 701a13) and (B1) 'I must make what is good' (DMA 701a17). Then (A3) 'I must march' gives us a necessary, but not sufficient, condition for the satisfaction of the imperative (A1). It is a necessary condition that I should march, if (A1) all men are to march and (A2) I am a man. But if I am the only one who marches, then this is not a *sufficient* condition for the satisfaction of (A1), "All men must march".

On the other hand (B3), 'I must make a house', gives us a sufficient, but not a necessary, condition for the satisfaction of the imperative (B1). If (B1) I must make what is good, and (B2) a house is a good thing, then my action expressing (B3) is sufficient to satisfy (B1). But there are indefinitely many other good things I could have made- daisy chains, romantic novels, box girder bridges- *any one* of which would have satisfied (B1), as it stands, equally well. So (B3) expresses an action the performance

of which is a sufficient but not a necessary condition for the satisfaction of (B1).

Hare's point is that Aristotle can't have it both ways. The necessary conditions of some moral imperatives can be satisfied by action; the sufficient conditions, of others; the necessary and sufficient conditions, of none. What makes the difference is where the universality comes in the first premiss.

This is a beautiful argument, which would surely succeed, *if* Aristotle's practical syllogisms, in canonical form, began with a universal imperative of either of the two types that Hare envisages. Unfortunately, however, as we have already seen, they do not. As David Wiggins puts it:

'There are no general principles or rules anyway... From the nature of the case the subject matter of the practical is indefinite and unforeseeable, and any supposed principle would have an indefinite number of exceptions. To understand what such exceptions would be and what makes them exceptions would be to understand something not reducible to rules or principles. The only metric we can impose on the subject matter of practice is the metric of the Lesbian rule [1137b27-32]... what Aristotle had in mind in NE VI was nothing remotely resembling what has been ascribed to him by his Kantian and other deontomaniac interpreters.'

(Wiggins 1980, p.231)

The validity of a practical syllogism will not be subject to this kind of lacuna if (in line with dA 434a 17-20, and as I have argued in (b-d)) its first premiss states simply that, for such a sort of person (namely, a good person), *this* sort of conduct is right and appropriate, and that *this* action here would be a case of that sort of conduct, and that I am that sort of person; so I should do it. This kind of inference seems difficult to fault on logical grounds, loosely phrased though it may (necessarily) be. Here, evidently, the analogy between practical and standard syllogism is in one way at its closest.

In fact, the difficulty and the interesting problem- to sum up- is not to formulate *valid* practical syllogisms, but *sound* ones: practical syllogisms which are not just valid, but true also. From a first premiss which states that, for such a sort of person, *this* sort of conduct is right and appropriate, plus the further premisses that *this* action here would be a case of that sort of conduct and that he is that sort of person (etc.), any fool can deduce that he ought to do the action in question. But

λεκτέον δ' ἴσως βουλευτὸν οὐχ ὑπὲρ οὗ βουλεύσασιτ' ἂν τις ἡλίθιος ἢ μαινόμενος, ἀλλ' ὑπὲρ ὧν ὁ νοῦν ἔχων (NE 1112a20-21).

We want to start our practical syllogisms from the kind of ἀρχαί, the kind of first premisses and second premisses, which the man of excellence would start *his* practical syllogisms from. And *this* is what is difficult and interesting, for the acquisition of these ἀρχαί is in itself the whole work of virtue.

5. Conclusion

In concluding this chapter, I bring to a close the main part of my exposition of Aristotle's theory of the voluntary. By now it should be plain what, on my account, Aristotle thinks voluntary action is, and how the nature of the voluntary is, for him, relevant to the problem of human freedom.

On Aristotle's view, we may say of voluntary action that it is action which is not compelled or done under duress; not done in ignorance; and not irrational. More informatively, we may also say a good deal about what an action must be like for it to satisfy these three conditions. It must originate within the agent, the agent being able to do otherwise. It must be action on well founded information and moral belief(s), where 'well founded' means that the agent who has the information and holds the moral belief(s) has, and has employed, the appropriate epistemological expertises for gathering information and formulating moral belief(s). And the action must be rationally explicable by means of a process of practical reasoning which is at least potentially, and preferably actually, occurrent in the agent.

If there are any actions which do satisfy these three conditions, then in Aristotle's view there is voluntary action. And if there is voluntary action, then there is free action; for to talk of free action could mean no more than to talk about voluntary action.

'Virtue is up to us, as likewise is vice. For in those cases where it is our choice (ἐφ' ἡμῶν) if we act, it will also be our choice if we do not act; in cases where consent is possible, so is refusal. Thus if it is up to us to act, where acting would be good, it is also up to us to refrain, where refraining would be bad; and *vice versa*. Now since it is up to us to do good or bad things, and likewise to refrain from doing them; and since how we act was the criterion of whether we are good or bad people; it follows that it is up to us whether we are worthwhile or worthless people. For the saying that 'No one's a willing scoundrel, nor yet unwillingly blest'- the latter part of this seems true, the former false. It is true that no one is unwillingly blessed; but wickedness is voluntary- unless we are to contradict what has just been said, and man is not to be called the origin and progenitor of his conduct, just as of his children. But if that seems right, and we cannot trace back our conduct to other sources than those within us, then what has its origins in us is up to us and voluntary.' (NE 1113b6-22)

Aristotle, then, holds that 'free action' is nothing else but voluntary action as he defines it: thus the NE III.1 account of the voluntary and involuntary is supposed to cover *all* the states which count as either. His theory of voluntary action is meant to be exhaustive as an account of free action. There *is* no free action which is not voluntary action in the sense which Aristotle gives that term.

But here there arises, for Aristotle's account, a question

which the phrase οὐδείς ἐκὼν πονηρὸς (1113b15, above) might seem to presage. What about the case of akrasia? How does this fit into Aristotle's theory of the voluntary? To that question, lastly, I turn in Ch.4.

to the
to the
to the

though
policy
harmful

1941

In 1941
of interest
there is a
strategic position
of importance
that which is
agreement which
he did not
action's general
and I think
quality, in 41, I will

The Varieties of Akrasia

1. How to Solve the Problem of Akrasia
2. How not to Solve the Problem of Akrasia: Aristotle's Account
3. The Varieties of Akrasia

'Those who restrain desire, do so because theirs is weak enough to be restrained; and the restrainer or reason usurps its place and governs the unwilling. & being restrained it by degrees becomes passive till it is only the shadow of desire.'

(William Blake, *The Marriage of Heaven and Hell*)

In this chapter, I will first try to show that, if my account of Aristotle's theory of the voluntary in Chs.1-3 is right, then there is a way of dealing with a problem which the phenomenon of akrasia presents for his theory (§1). I shall develop this line of argument by reference to the modern debate about akrasia, and then relate this debate to my exegesis of Aristotle. The line of argument which I develop was, in principle, open to Aristotle, but he did not use it. Then I will show how very different was Aristotle's actual treatment of akrasia in NE VII; and I will say what I think are the inadequacies of this account of akrasia. Finally, in §3, I will point to some instances of the great

variety of kinds of akrasia or akrasia-like conditions which I believe there are. This chapter aims to show that, *vis à vis* the Aristotelian account of voluntary and free action which I have presented in Chs.1-3, the existence of something like akrasia not only is no threat, but provides an interesting supplement to that account.

1. How to Solve the Problem of Akrasia

At a time t someone, N , is confronted with two courses of action A and B . N believes N can do either A or B at t , but not both; and N judges at t that it would be better for N to do A and not B at t . Nonetheless N voluntarily does B and not A at t .

This phenomenon has had various English names in recent discussion: 'backsliding' or 'hypocrisy' (Hare 1952, 1963), 'weakness of will' (Davidson 1980, Charlton 1988), 'incontinence' (Davidson 1980). But these English names beg important questions, so I shall stick with 'akrasia', the Greek name that Aristotle himself uses, and its derivative 'akrates' to describe the agent who suffers from akrasia.

Does akrasia occur? Some philosophers, more or less following Aristotle (NE 1145b28), have taken it to be a rather obvious truth of experience that they do. Other philosophers, more or less following Socrates (*Protagoras* 355c), have taken it to be a rather obvious truth of logic that they don't. For the former

school, the problem of akrasia is the problem of explaining *how* akrasia *is* possible. For the latter school, it is the problem of explaining *why* akrasia *isn't* possible.

As Justin Gosling has recently observed (Gosling 1990, p.196-7), modern discussions of akrasia have mainly taken their cue either from Richard Hare or from Donald Davidson, whose celebrated treatments of akrasia display this contrast rather well (Hare (1952) and Hare (1963); Davidson, 'How is Weakness of the Will Possible?', in Davidson, 1980, p.21-42). Re-examination of the alternatives which Hare and Davidson offer us suggests a way of reconciling the attractive, but apparently incompatible, intuitions on which they work. It also suggests a way of doing justice to a third intuition.

Hare's explanation why akrasia isn't possible, at least not in the full-blooded sense in which many think it is possible, follows directly from his account of what moral beliefs are. As every undergraduate knows, on Hare's account any belief, to count as a moral belief, must display two features, *prescriptivity* and *universalisability*. The idea of universalisability is this: If I make judgement J1 in situation S1, then J1 is a universalised judgement if and only if, when I make J1, I could commit myself to making the relevantly similar judgements J2-Jn in all relevantly similar situations S2-Sn. The idea of prescriptivity is that 'Moral judgements, in their central use, have it as their function to guide conduct' (Hare, 1963, p.70).

Hare thinks that these two requirements are together necessary and sufficient to characterise moral beliefs. To require of moral beliefs only that they should be universalisable, and not also prescriptive, would let in as a moral belief, e.g., 'This thing is red' (Hare 1963, p.11). To require only that they should be prescriptive, and not also universalisable, would let in as a moral belief, e.g., 'I will jump/ want to jump this queue'.

So Hare can dismiss the possibility of full-blooded akrasia like this. If akrasia means that I act against a moral belief in Hare's sense, then Hare can simply deny that any belief against which I in fact acted *could* be a moral belief. It must be 'off colour' in 'one of the many ways that are possible' (Hare 1963, p.68). For if a belief of mine is both fully universalisable and fully prescriptive, then necessarily, I will act on it unless something prevents me. For Hare it is a point of logic, a point about the use of words, that any case where I appear to have a fully formed moral belief, but act against it, must be a delusion. Akrasia is not that case, but a kind or kinds of case which *looks like it*:

'So difficult is it, in fact- so great is the strain between prescriptivity and universalisability in certain situations- that something has to give: and this is the explanation of the phenomenon of moral weakness.'

(Hare 1963, p.73)

Davidson, on the other hand, means to explain how akrasia is possible. Though he does not put it quite like this, his method is, in effect, to distinguish three (not two) kinds of practical judgements: (i) *sans phrase* judgements, (ii) *prima facie* judgements, and (iii) 'all things considered' judgements.

(i) *Sans phrase* judgements are practical judgements which are indefeasible, final, and unconditional; an example would be the unqualified 'N should do A at t and not B', where this decision does not fail to lead to action. For it is *sans phrase* judgements which stand in direct causal relation to actions.

(ii) *Prima facie* judgements, on the other hand, are qualified or, as Davidson says, 'relational' practical judgements. Their persuasive force depends on their relation to evaluative generalisations about actions which can only provisionally be assumed to hold good (Davidson 1980, p.39). Thus, e.g., the judgement that '*Prima facie*, N should do A at t and not B' would depend on the generalisation that actions of A's action-type are as a rule preferable to actions of B's action-type. *Prima facie* practical judgements, then, lack the practical cogency of *sans phrase* judgements.

(iii) Davidson's third kind of practical judgement is the 'all things considered' judgement, the judgement based on all (or is it: the majority of?) the evidence available.

Now it might seem that the judgement '*All things considered*,

N should do A at t and not B' can simply be equated with the *sans phrase* judgement 'N should do A at t and not B'. But the nub of Davidson's argument for the possibility of akrasia is to deny this equation. Despite some equivocation on the point (Davidson 1980, p.40, para.2), it seems fairly clear that his central claim is that (iii) 'all things considered' judgements are equivalent, not to (i) *sans phrase* judgements, but to (ii) *prima facie* judgements. Hence Davidson can say that the 'all things considered' judgement, against which the akrates acts, is not a *sans phrase* judgement, but a *prima facie* judgement. For Davidson, this counts as establishing the possibility of akrasia because there is no logical contradiction between the akrates' two judgements. For an 'all things considered' judgement, being logically equivalent only to a *prima facie* judgement, cannot conflict with a *sans phrase* judgement.

Hare and Davidson seem to me to represent two extremes, between which I want to find a mean. For in important respects I find myself agreeing and disagreeing with both of them.

Hare's central intuition that there is something logically wrong with the very idea of akrasia is one which many have found suspiciously aprioristic. So Lemmon (1962):

'Hare should not have defined value judgements in such a way that sincere assent to them entails an imperative leading to action, since in quite normal senses of the word a man precisely does assent to a value judgement sincerely and still fails to act

accordingly, in the situation of akrasia.'

(Lemmon 1962, p.144)

Contra Lemmon, I think Hare was on to something here. There is a sense in which there *is* something logically incoherent about the very notion of akrasia. My disagreement with Hare is only about where this incoherence is to be located. For I agree with Davidson that what is in question in the problem of akrasia is not the nature of *moral* belief as such, but the nature of *practical rationality*: 'Incontinence is not essentially a problem in moral philosophy, but a problem in the philosophy of action' (Davidson 1980, p.30, n.14). (This is a point which might also be suggested by a reading of Philippa Foot; v., in particular, her essay 'Are Moral Considerations Overriding?' in Foot (1977).)

On the other hand, Davidson seems to argue against (Hare's?) idea that akrasia is a logical impossibility *via* an argument that the practical reasoning which leads to akrasia involves no logical contradiction. Thus Davidson seems happy to admit that the akrates acts against 'rationality' (in Davidson's sense of that word):

'Why would anyone ever perform an action when he thought that, all things considered, another action would be better? ...What is special in [akrasia] is that the actor cannot understand himself: he recognises, in his own intentional behaviour, something essentially surd.'

(Davidson 1980, p.42)

But what Davidson will *not* admit is the idea that the akrates' practical deliberation might involve a formal contradiction. No doubt akrasia is unreasonable behaviour, but (Davidson thinks) it cannot, quite, actually be illogical behaviour, in the sense of being based upon a logical mistake. Davidson considers his job done once he has shown that

'the logical difficulty has vanished because a judgement that a is better than b, all things considered, is a relational or [*prima facie*] judgement, and so cannot conflict logically with any unconditional judgement'.

(Davidson 1980, p.39, my emphasis)

But is showing that akrasia is not based on a logical mistake really the same as showing that akrasia is not logically impossible? After all, anyone who says sincerely that three and three make eight is making a logical mistake, but this does not show that it is logically impossible to make that mistake. *Contra* Davidson, it seems an attractive and plausible suggestion to say that what we are trying to describe, when we deal with akrasia, is a logically confused process of practical reasoning. But of course it would not follow that any description of that process was bound to be confused.

In fact, I will not argue here that (full blooded) akrasia is

based on a logically confused process of practical reasoning. But I will argue one quite similar point, viz. that one important species of akrasia involves a logical incoherence, not between different parts of the same practical reasoning process, but between the practical reasoning process which actually occurs and the practical reasoning process which we might want to infer on the basis of the akrates' actual behaviour. It is just because this is so that I am also inclined to agree with Hare that there is at least one important kind of akrasia (though not the same type) which cannot satisfy the description of akrasia with which this chapter begins without logical inconsistency. My case depends on four points, which between them should now mesh what I have just said about Hare and Davidson with the account of Aristotle's theory of action which I presented in Chs.1-3.

1. I will say that there are such things as *practical imperatives*. These, very generally, are motivations to act of any and every conscious type.

2. Practical imperatives are of two sorts: some are *conditional* and others are *unconditional*. Conditional practical imperatives (CPIs) have the form 'In view of some reason(s) it would be good for me to do x'. Unconditional practical imperatives (UPIs) have the form 'Taking into account all conditional practical imperatives, it is best for me to do x'.

The point of this distinction is that any CPI formulated at time *t* is *logically overridden* by any UPI formulated at *t*. For

any CPI formulated at a given time is 'contained in' any UPI formulated at that time: the UPI is formed precisely by the aggregative consideration of all CPIs. So the agent might reason, in a very simple case, like this:

- (a) I want to go to the pub tonight.
- (b) But I said I would visit my aunt tonight, so I ought to do so.
- (c) One should keep one's promises unless there is good reason not to; and there isn't.
- (d) So, all things (i.e., (a-c)) considered, it is best for me to visit my aunt tonight.

Here (a) mentions a desire, (b) an obligation, (c) a moral principle and its application. But none of (a-c) is any more than a CPI. The only UPI here is (d). My point is that (d) is arrived at simply by taking into account all the factors conceived as relevant, which in this case means (a-c).

(This piece of practical reasoning, which goes to form the UPI (d), is not, of course, a practical syllogism. It is a practical *induction*, a dialectical process which issues in the definition, or specification, of (d) as the first premiss of a subsequent practical syllogism. The unconditionality of (d) is then 'transmitted', by that practical syllogism, to its conclusion: if the practical syllogism says that getting on the bus is a necessary means to fulfilling the UPI (d), then 'I must get on the bus' will itself become a subsidiary UPI for the agent.)

3. Hence only some practical imperatives, viz. the UPIs, are practically cogent, have indefeasible persuasive or action-motivating force. While all practical imperatives provide the agent with some sort of reason to act, not all practical imperatives, but only UPIs, provide the agent with *necessary and sufficient* reason to act. This means that only UPIs are such as to be suitable to enter into an account of the causation of voluntary action.

(Consider the contrast between 'natural' or 'physical' causation, and the causation of voluntary action. In natural causation, a cause C brings about its effect E because C *actually* does provide necessary and sufficient reason for E to occur. In the causation of voluntary action, by contrast, the agent A's reason R to do X brings it about that A does X because R *seems to* A to provide necessary and sufficient reason for A to do X. Note the subjective and intensional nature of this account.)

Thus if some agent who acts formulates only a CPI, then something is missing from the explanation of his action as a voluntary action. For, while a CPI does provide some sort of motivation to action, it *alone* cannot have provided the agent with a necessary and sufficient motivation to act. For that we need to postulate that there was also a UPI, even if this UPI differs from the CPI in no respect except that of being unconditional.

4. To summarise the points made about the conditions of

voluntary action in Chs.1-3: to understand a piece of behaviour as a voluntary action involves three assumptions about that behaviour: (i) that the agent was *not compelled* to do what she did; (ii) that she acted *without ignorance* of any important relevant consideration; and (iii) that her behaviour was *rationaly explicable*, by which I mean that we have to assume that it was done as a result of a piece of practical reasoning which led to the formulation of what I have called a UPI in the way already described in the last chapter.

Now, given these four points, I can show in what sense akrasia involves a logical incoherence between thought and act; and in what sense the concept of akrasia is itself a logical incoherence. Let us ask: what are the *necessary and sufficient conditions* for a case of akrasia? I suggest that they are six in number. Three of them together stipulate that the akratic behaviour be a *voluntary action*. Three of them concern the akrates' beliefs at the time of action, and together stipulate that the akrates must knowingly do what she simultaneously believes it would be better not to do. The former three conditions are, of course, based on the three conditions of voluntary action which I have argued for. The latter three conditions of akrasia bring into play the contrast between UPIs and CPI s which I have just established. These then will be the severally necessary and jointly sufficient conditions for full akrasia:

Beliefs: For X's doing A at the time t1 to be a case of akrasia, X must do A at t1, and consciously hold at t1 beliefs which could be expressed in these words:

- C1. 'Unconditionally, A is not to be done by X at t1.'
- C2. 'Conditionally, A is to be done by X at t1.'
- C3. '"Unconditionally, A is not to be done by X at t1" overrides "Conditionally, A is to be done by X at t1".'

Voluntariness: For X's doing A at t1 to be a case of akrasia, X must also do A at t1 in these three ways:

- C4. without compulsion of any external or internal kind,
- C5. without important relevant ignorance, and
- C6. in a way that is not rationally inexplicable.

Taking these to be the necessary and sufficient conditions of full blooded akrasia allows me to draw the following three conclusions.

(i) There is a kind of akrasia, namely full blooded akrasia itself, the very idea of which does indeed, as Hare says, involve a logical confusion- though not the confusion Hare had in mind. The point is that C1-3 are logically incompatible with C6. If C1-3 are satisfied, then it follows that C6 is not satisfied, and *vice versa*.

How so? Consider again the notion of the rational

explicability of a voluntary action. Explaining a voluntary action (I have suggested) means giving an account of the causation of the behaviour which it involves, in terms of the agent's seeing himself as having necessary and sufficient reason to do what he does. The presence in the agent's thinking of such (to the agent) necessary and sufficient reasons for action is the essence of any action's rational explicability. Now a necessary and sufficient reason for action can only be provided by an *unconditional practical imperative*. Hence properly to explain any piece of behaviour as a voluntary action is to cite the UPI which provided the necessary and sufficient reason for that action's performance.

But consider the fully akratic action, A, as just characterised by my C1-6. *Upon what unconditional practical imperative is A performed by X at t1?* There is (of course) only one UPI which X, as described above, entertains at t1. But not only is this not a UPI to *do A at t1*; it is actually a UPI to *do not-A at t1*!

By the canons of rational explicability which I have suggested here, the only voluntary action at t1 of which the beliefs given in C1-3 could be explanatory is any voluntary action or abstention from action which counts as *not doing A at t1*. But, in the case of full akrasia, such an action is precisely what we don't have. X does the practical reasoning which, for the purposes of rational explicability, ought to go with voluntarily not doing A at t1; and yet the action which X does at t1- is A.

it follows that A, so performed, cannot satisfy C6. Hence A, so performed, is not a voluntary action. And- since full blooded akrasia is supposed to be voluntary action- A is not a case of full blooded akrasia either.

Thus the sense in which the very notion of akrasia is incoherent is nothing to do with the nature of moral judgement. It is rather the confusion of thinking that a piece of behaviour could be rationally inexplicable in the sense I have defined, and yet still also be a fully voluntary action. A case of full blooded akrasia, satisfying all of C1-6, would have to be like that. But it is impossible for any case to be like that because unless it is rationally explicable, a piece of behaviour does not qualify as a voluntary action at all, but remains merely unexplained behaviour. About the causation of such behaviour we do not know what we must know to be able to count it as a voluntary action.

(ii) There is also a kind of akrasia- though it is not 'full blooded' akrasia, like the kind just discussed which is of a logically incoherent form- which does indeed involve what Davidson calls 'something surd', and even an actual logical inconsistency. The inconsistency in question is not so much *in* the agent's practical reasoning, as *between* the agent's actual practical reasoning and the practical reasoning that ought to go, but does not, with the agent's action. Take the case where C1-5 are satisfied, though C6 is not (for brevity, call this case (12345)). The agent's thinking leads, or ought to lead, the agent

not to do the akratic action; yet the agent does the akratic action. There is nothing logically impossible about the occurrence of this sort of akrasia, even though it seems to involve something very like a logical inconsistency. Such a case is simply *puzzling*; one is inclined to say of it, with Davidson, that 'In the case of incontinence the attempt to read reason into behaviour is *necessarily* subject to a degree of frustration' (Davidson 1980, p.42).

(iii) My third conclusion underlies the analysis which has led to the first two conclusions; and it answers to a third intuition which Hare and Davidson share (as also does Gosling). This is that there are many more than one type of cases which look like akrasia:

'There are many different methods of backsliding without appearing to.'

(Hare 1963, p.76)

'We are dying to say: remember the enormous variety of ways a man can believe or hold something, know it, or want something, or be afraid of it, or do something... These half-states and contradictory states are common, and full of interest to the philosopher.'

(Davidson 1980, p.28)

'In practice many different kinds of behaviour are commonly described as showing weakness [of the will] which do not feature, or not largely, in traditional accounts. Consideration of some of them suggests a criss-cross of overlapping problems...'

(Gosling 1990, p.167)

How many kinds of behaviour? And how are we to classify them? If there are six severally necessary and jointly sufficient conditions of full blooded akrasia, then perhaps the way to classify the varieties of akrasia is by means of a statistical array like the one to be found at the end of this chapter. Since there are 2^6 (=64) ways in which any six conditions can be either satisfied or not, there will be sixty-four different types of case answering to all the varieties of partial, full blooded and non-akrasia which exist, might exist, or demonstrably cannot exist. Drawing up this array may make it clear which types of akrasia logically can not occur; what the range of possible partial akrasias is; and what kinds of cases are not akrasias at all.

However, consideration of these will be deferred until §3; what I want to do next, in §2, is consider Aristotle's account of akrasia.

2. How not to Solve the Problem of Akrasia: Aristotle's Account

If full blooded akrasia existed, it would pose a threat to Aristotle's doctrine of προαίρεσις. As we have seen, Aristotle's account of the origination of action is (leaving aside variations which do not matter here) that practical reasoning generates a προαίρεσις, which we might perhaps think of (in this context) as a decision to act; and the προαίρεσις in turn generates a voluntary action. Now akrasia, apparently, is voluntary action (EE 1223a37-b3); yet akrasia is 'against the προαίρεσις' (NE 1148a18). Aristotle argues that voluntary action is definable as action which is caused by προαίρεσις, from which it ought to follow that there is no voluntary action which is *not* caused by προαίρεσις. The threat is that full blooded akrasia may provide a counter example to this claim.

Aristotle can deal with this threat by employing something like my distinction between a CPI and a UPI to deny that anything which could be called full blooded akrasia ever occurs. He can equate προαίρεσις with the UPI, and the three kinds of ὀρεξις (and other kinds of conditional motivations, such as those provided by νοῦς, NE 1139a36) with the CPI. Then suppose that a piece of behaviour B occurs in which the agent acts contrary to προαίρεσις. Expressed in my terms, what Aristotle can say is that, because B is not action on any UPI, and hence is not rationally explicable, B is not a voluntary action at all; hence B is not a case of akrasia either.

Aristotle does come quite close to equating προαίρεσις with what I call the UPI, and the three kinds of ὀρεξεις and other kinds of conditional motivations with the CPI. Consider Aristotle's account of the nature of προαίρεσις (and βούλευσις) in NE III.2-3. The central point of this is (Ch.3, §1) that προαίρεσις is neither belief nor appetite. We can now see that the reason why προαίρεσις is not just belief is because belief is, in itself, at most only a conditionally motivating factor- a CPI, in other words; but προαίρεσις is *unconditionally motivating*, because a προαίρεσις expresses a UPI. Again, προαίρεσις is not just appetite, because the three forms of appetite, although they are motivating factors, are not *unconditional* motivations- but προαίρεσις *is*, because it expresses a UPI.

So Aristotle repeatedly stresses that (unlike either δόξα or any sort of ὀρεξις) προαίρεσις, or sometimes βούλευσις, issues directly in the action which it causes:

οἱ δὲ λέγοντες [τὴν προαίρεσιν] ἐπιθυμίαν ἢ θυμὸν ἢ βούλησιν ἢ τινα δόξαν οὐκ εὐόκασιν ὀρθῶς λέγειν... (NE 1111b11-12)

φαίνεται δ' ἡ μὲν ζήτησις οὐ πᾶσα εἶναι βούλευσις, οἷον αἱ μαθηματικά, ἡ δὲ βούλευσις πᾶσα ζήτησις- καὶ τὸ ἔσχατον ἐν τῇ ἀναλύσει πρῶτον εἶναι ἐν τῇ γενέσει. (NE 1112b23-25)

προαίρεσις, then, is both necessary and sufficient to motivate the action to which it leads. In the terms which I am

arguing for, we might say that a complete deliberation gives an unconditional motivation to act— a UPI. It gives such a motivation because (going forwards) a deliberation is incomplete *until* it actually causes an action (1112b25); and also because (going backwards) a *complete* deliberation involves not only CPIs but also a UPI.

So, if the real problem about *akrasia* which Aristotle's account must deal with is the question: 'Can there be voluntary action on what Aristotle classes as a conditional motivation and against what he classes as an unconditional motivation?', Aristotle could have had a neat response to this question. He could have allowed for the possibility of behaviour which goes against an unconditional motivation, and is apparently action on a conditional motivation. But he could also have argued that such behaviour is not explicable by reference to the agent's actual or potential deliberations. For the combination of an unconditional motivation to do X and a conditional motivation to do not-X yields an unconditional motivation to do X; but the *akrates* is the person who, in this situation, does not-X. Hence there is nothing in the agent's actual or potential deliberations which explains his actions. Hence his behaviour is not rational; and hence in turn his behaviour is not voluntary action.

But this is not, in fact, anything like the account of *akrasia* which Aristotle *does* give us in NE VII. Aristotle's account of *akrasia* in NE VII is dominated by the drawing of distinctions about different senses in which an agent can be said

to 'have the knowledge' of something when acting. These distinctions are brought out in response to the Socratic challenge recognised at 1145b23-26: 'It would be strange if, when knowledge was present in the agent, something else could conquer that knowledge and drag it about like a slave.' Aristotle thinks he has met this challenge by 1147b15: 'It seems that what Socrates sought has come about...'

To treat 'the problem of akrasia' in these *Socratic* terms is foreign to Aristotle's method elsewhere, and largely irrelevant to the real problem raised for his account by akrasia. Aristotle, I suspect, has taken his eye off the ball here. He is not following through on his own agenda; his attention has turned to a rather different debate. He produces a free-standing, set-piece response to a Socratic question, rather than an integrated series of answers to Aristotelian questions. This does not prevent Aristotle's account from giving us much interesting material, which I shall now examine.

Given the way Aristotle's account of akrasia stands (most of it) between the raising of a Socratic question and that same question's laying to rest, we ought to find a certain unity in the account. And so we do. Nonetheless, we may discern six separate phases in Aristotle's treatment of akrasia, marked out by introducing-words almost as if the phases were meant as separate accounts:

1146b31 ἄλλ' ἐπεὶ διχῶς λέγομεν...

1146b36 ἔτι ἐπεὶ δύο τρόποι...

1147a4 διαφέρει δὲ καὶ τὸ καθόλου...

1147a11 ἔτι τὸ ἔχειν τὴν ἐπιστήμην ἄλλον τρόπον τῶν νῦν ῥηθέντων
ὑπάρχει τοῖς ἀνθρώποις...

1147a24 ἔτι καὶ ὧδε...

1150b19 ἀκρασίας δὲ τὸ μὲν προπέτεια τὸ δ' ἀσθένεια...

These are the six phases so marked:

1. 1146b31-36: the potential/ actual knowledge distinction. (§2a)
2. 1146b36-1147a4: application of this. (§2b)
3. 1147a4-10: the 'syllogism breakdown'. (§2c)
4. 1147a11-24: the conscious/ unconscious knowledge distinction.
(§2d)
5. 1147a25-b18: the combined account. (§2f)
6. 1150b19-28: the impetuosity/ weakness distinction. (§2e)

Five of these (given between 1145b24 and 1147b15) have an obvious unity. The relation of the sixth phase to the other five is less clear. And, if Aristotle meant even these first five phases to relate only to the Socratic question, he did not succeed in keeping the issue that clear-cut.

The most complex of these, 1147a24-b18, in which he seems to be trying to bring into play together all the ideas he has raised in phases 1-4, is apparently Aristotle's most considered opinion on akrasia. The sixth account looks like it was interpolated

later by Aristotle— the afterthought of someone uncomfortably aware that he has not really settled the matter? As the fifth account is an attempt to combine elements of the first four (and other elements too), it is probably best left till last. The place to start is with the first account, for this seems to be involved in all the others except the sixth.

2a. Potential/ Actual Knowledge

At 1146b33-35, Aristotle distinguishes two ways in which someone can be said to 'have knowledge' (ἐπίστασθαι):

'Someone who has knowledge, but is not using it, is said to know; so is someone who is using that knowledge. So the case where a [wrongdoer] has knowledge [that what he does is wrong] but is not considering it (θεωροῦντα) will be different from the case where a man is considering what he ought not to do. It would indeed (γὰρ) seem strange [if he did wrong when he was considering that knowledge]; but not if he was not considering it.'

I have already invoked EE 1225b11-12 as evidence for a potential/ actual knowledge distinction in Aristotle (Ch.1, §3b; Ch.3, §3). That distinction is this: I *actually know* P at t1 if I state, to myself or someone else, that P at t1; I *potentially know* P at t1 if I *could* state, to myself or someone else, that P at t1, but (for one reason or another) do not do so (and I do not

know that P at t1 if I *neither do nor could* state, to myself or someone else, that P at t1). Likewise (I said), I formulate, at t1, a practical syllogism S *actually* if I state S, to myself or someone else, at t1; *potentially* if I *could* have stated S, to myself or someone else, at t1; and *not at all* if I *neither do nor could* state, to myself or someone else, S at t1. Hence, my action A at t1, given that S is explicable by the practical syllogism S, is *actually* rational if I state S, to myself or someone else, at t1; *potentially* if I *could* have stated S, to myself or someone else, at t1; and *not at all* if I *neither do nor could* state, to myself or someone else, S at t1.

This state of affairs makes it impossible to use the actual/potential knowledge distinction to explain akrasia. The point which Aristotle wants to make about potential knowledge here is that the action which should be performed is not performed *because* the agent's knowledge is only potential. Evidently this could happen. When, for example, I forget to meet John at the station for an afternoon's train-spotting, I *could* make the statement 'I said I'd meet John at Waverley at 2 p.m.', if I was suitably prompted ('Where are you meant to be at 2?'). So I do know the fact, considered simply as a piece of knowledge, potentially but not actually.

But, as far as the explanation of action goes, such knowledge does not count even as potential knowledge. Potential knowledge *which explains action* is knowledge which I *act upon*, although I make no actual statement of the knowledge (but could if

prompted). Thus, if I went to the newsagent's at 1.45 and went straight on, without thinking, from there to Waverley, this action would display potential knowledge which was motivating the action: although I have not formulated the thought 'I said I'd meet John at Waverley at 2 p.m.', this is what makes me go to Waverley. Potential knowledge which explains action is the kind of thing which one *could* mention- if one were to formulate it- in answer to the question 'Why are you doing that?'. But if there is no action to which some potential knowledge K is relevant as explanation, then, as far as the explanation of action goes, I do not potentially know K; I am actually ignorant of K. Thus any case of akrasia which involves such 'forgetfulness' will not be a case of full akrasia (123456), but of akrasia mitigated by ignorance: that is, at best, (1234).

2b. The Application of this Distinction

Aristotle, nonetheless, applies this idea of akrasia as involving 'forgetfulness' at 1146b36-1147a4:

'Again, since there are two forms of the premisses [sc., both universal and particular], nothing prevents someone who has both from acting contrary to knowledge when his knowledge of the universal is actual but his knowledge of the particular is only potential. For action has to do with particular things.'

If someone's knowledge of a particular premiss fails to get

actualised, and remains only potential (in a form that has no bearing on the explanation of the action), then, as we have just seen, there is no good reason to say that the agent really knows the premiss. If it is forgotten information, then it is not (in the relevant sense) knowledge at all.

(Why, incidentally, must it be the knowledge of the particular premiss which fails to get actualised? Couldn't the knowledge of the 'universal' fail to get actualised too? Because, as we saw in Ch.2, the more general, and the more explicitly moral, a principle seems, the less easy it is to envisage it being forgotten.)

2c. 'Syllogism Breakdown'

Aristotle keeps us guessing by giving us a second, more complicated, example of his actual/ potential knowledge distinction:

'There is also a distinction as to the universal. For one universal relates to the agent and one to the thing [cp. NE 1105b309 ff., distinguishing the mean of the thing from the mean relative to us]. Thus, the agent may know that dry food is good for every human, and that he himself is a human, or that foodstuff X is dry food. But that *this* is foodstuff X- either he may not be exercising this knowledge, or *he may not have it at all.*' [My italics]

Whatever are we to make of this strange passage? In the first place, Aristotle is snubbing grievously a distinction which, only fifteen lines before, was getting VIP treatment- although, I have argued, it does not deserve it. I mean the distinction between having knowledge (but only potentially), and not having it at all. My italics are meant to point up the alarming way in which Aristotle glides from one to the other here. But this is no small difference, as Aristotle's own arguments about potential and actual knowledge are meant to suggest. If any piece of particular knowledge simply isn't there, as opposed to being *in potentia* only, then something vital is missing from the description of the agent as an akrates. For the agent is ignorant of something which is vital to the syllogism. And (as we have seen) to be ignorant is to be less than fully voluntary; and to be less than fully voluntary is to be less than fully akratic.

Secondly, if there are two 'universals' here, Aristotle does not tell us what the second 'universal' is, which presumably is the one relating to the agent, not the thing. 'Dry food is good for every human' is evidently the 'universal of the thing'. 'He himself is a human', 'Foodstuff X is dry food', and 'This is foodstuff X' are all, surely, particular premisses. So what 'universal' is missing? And what is its relevance to what Aristotle is saying here, anyway?

The missing 'universal' can be supplied. It seems that Aristotle wants us to picture a practical sorites which, made completely explicit, goes like this:

- A. Dry food is good for humans to eat.
- B. I am a human.
- C. Dry food is good for me to eat.

- D. Dry food is good for me to eat.
- E. Foodstuff X is dry food.
- F. Foodstuff X is good for me to eat.

- G. Foodstuff X is good for me to eat.
- H. *This* is foodstuff X.
- I. [He eats the bit of X.]

Here 'the universal of the agent' is, apparently, (D). (There is also (G): Aristotle must be conflating the second and third syllogisms of this sorites if he thinks that (D) is the only 'universal of the agent'.) But now what point is Aristotle making by drawing our attention to this 'second' universal? For the premiss that he suggests is the mischief maker in a practical sorites like this is, as we might expect, *none* of the universal premisses (A), (D), and (G). It is one of the particular premisses (presumably either (E) or (H)). What Aristotle really means to do, by drawing our attention to the distinction between the two (or rather three) *universals*, is to make a point about the two (or rather three) *syllogisms* which compose the sorites.

Then what does Aristotle mean when he says *this*?-

'Now indeed there will be a very great difference between

these types (τρόπους), so that it will seem to be nothing absurd if [the akrates] knows in this way (οὕτω); but amazing if he should know otherwise (ἄλλως).' (NE 1147a8-10)

What 'types' is Aristotle distinguishing here? (i) Types of premiss, as particular and universal (he uses τρόποι elsewhere in this passage only at 1147a1, apparently of *that* distinction)? (ii) Types of particular premiss, as 'particular of the thing' and 'particular of the agent'? (iii) Types of knowledge, as potential and actual? (iv) Or some combination of these types?

If (i), then Aristotle is restating baldly his preference that the universal premiss should not be the one that gets obscured in any way. (ii) would fit most neatly into the context; but then is the agent actually *ignorant* either of (E) or of (H)? (ii) and (iii) combined- so that the agent is (in Aristotle's sense) only potentially aware either of (E) or of (H)- seem to offer the best way of making sense of what Aristotle says here. But taking any of the four options will leave us with the puzzle of why it is supposed to be less surprising that the agent should be either actually or effectively ignorant in one way than another. Aristotle gives us little help with this.

Probably, the passage means this:

'In view of the different types [of particular knowledge] there are, it is not surprising that the akrates only has [potential knowledge of one of the particular premisses (E) or

(H)]- but it would be very surprising if he had [actual knowledge].'

But why should reflection on the variety of the types of particular knowledge lead us to expect the akrates only to have potential knowledge? Are we to suppose that the akrates becomes *confused* between them (in which case, again, he is not an akrates, but is ignorant)? Aristotle remains uninformative on this.

2d. The Conscious/ Unconscious Knowledge Distinction

Aristotle has not finished making fine distinctions. Not only (he thinks) is there a type, potential knowledge, opposed to actual knowledge; there are two types of potential knowledge.

'In the state of having knowledge but not exercising it [= potential knowledge], we can see a distinction as to the disposition (τὴν ἔξιλον), so that a person can as it were both have [potential] knowledge and not have it- like a sleeper or a madman or a drunkard. But persons under the influence of emotions (πάθεισιν) are like this... It is clear then that we can say that the akrateis are like this. The fact that they speak the words which come from knowledge means nothing... akrateis speak like actors playing parts.' (NE 1147a11-24)

Aristotle is working on reducing the meaningfulness of saying that 'akrateis have knowledge' till it has come almost to the

vanishing point- but is not quite there; except when he forgets himself and throws in talk about akrateis as not having certain kinds of knowledge at all. I have already commented on the irrelevance and fruitlessness, for Aristotle of all people, of this procedure as a way of dealing with akrasia. For if some relevant knowledge does not enter into the explanation of an action, then the agent acts in ignorance- irrespective of whether she *could* supply the information in question if prompted.

Still, what is the sense of this distinction which, for want of a better name, I call the 'conscious/ unconscious knowledge distinction'? Since Aristotle gives us no particular hints about what it means, apart from his examples, we have to shift for ourselves. One would like to say that the point is that sleepers, madmen and drunkards are under a *double counterfactual* where the possessors of ordinary (only) potential knowledge are under a *single* one. That is, a possessor of ordinary only-potential knowledge is in a state where this proposition is true of him: He would normally have been able to tell you the answer to your question if he were asked; but isn't so able for the duration of his akrasia. Whereas a possessor of unconscious only potential knowledge is in a state where *this* proposition is true of him: He would normally have been in the state of the possessor of ordinary only potential knowledge; but, for the duration of his akrasia, isn't in that state, even.

This suggestion is at least partly supported by the examples. Drunkards and madmen are afflicted by kinds of abnormal mentality

that put them at a double disadvantage when it comes to avoiding akrasia. In the first place, they might be akratic anyway, when not also mad or drunk. In the second place, madness and drunkenness certainly do not help us to remain enkratic. Madmen and drunkards will also provide good examples of knowledge that appears to be being 'used', in Aristotle's sense, but in fact isn't being: for the causal connections between stimulus and response, and information and knowledge, are all wrong in their cases. The drunkard who repeats the verses of Empedocles (1147b13) is indeed producing 'the words that come from knowledge' (1147a29), at least if you take Empedocles' verses to contain knowledge, but producing them fortuitously. However, this suggestion of a double counterfactual remains only a conjecture, and nothing has been done about the central problem: the apparent confusion between potential but action-guiding and potential but action-irrelevant knowledge.

2e. The Impetuosity/ Weakness Distinction

'Some akrasia is impetuosity (προπέτεια), other akrasia is weakness (ἀσθένεια). Some [the weak] deliberate, but through emotion (διὰ τὸ πάθος) do not remain in (οὐκ ἐμμένουσιν) the [resolves] which they have deliberated. Others [the impetuous] are led by emotion (ἄγονται ὑπὸ τοῦ πάθους) because they have *not* deliberated... The excitable [ἐκστατικοὶ, lit., 'the jumpers-out'] are better than [the weak], those who have reason (λόγον) but do not remain in it. For the [weak] succumb to

smaller temptations, and they have previously deliberated, unlike the other sort.' (1150b19-22, 1151a1-5)

Aristotle here seems to suggest two ways of accounting for akrasia. Akrasia A, impetuosity, occurs when an agent seems to act voluntarily against an unconditional motivation which she never actually formulates. Akrasia B, weakness, occurs when an agent seems to act voluntarily against an unconditional motivation which she does actually formulate, but does not act upon.

This fifth distinction, as I say, looks very like an afterthought. The focus of the rest of the discussion is on the status of the akrates with regard to *knowledge* and *ignorance*; but this distinction focuses on the akrates' status with regard to *appetition* and *compulsion*. Is it, then, any more successful as an account of akrasia? I think not. Neither of these 'akrasias', A or B, is in fact a genuine case of akrasia.

If (A) an agent seems to act voluntarily against an unconditional motivation which she never *actually* formulates, then there are two possibilities. Either (i) she formulates it *potentially*, in which case her behaviour will be irrational and so, as it does not satisfy condition (6), will not in fact be voluntary action. Or else (ii) she does not formulate it at all, in which case there is no sense in which she holds the belief mentioned in condition (1).

(1) But if (B) an agent seems to act voluntarily against an unconditional motivation which she does actually formulate, then, again, there are two possibilities. Either (i), if that motivation *is* unconditional, then her 'action' against it involves her holding the beliefs which satisfy my conditions (1-3); hence her behaviour, while it is a form of partial akrasia (12345), is not voluntary action because it is not rational. Or else (ii) that motivation is *not* unconditional, in which case her behaviour is either irrational, and therefore not full akrasia, or else is voluntary action (on some other unconditional motivation), but not any sort of akrasia.

2f. The Combined Account (1147a24-b18)

This is no greater than the sum of its parts, which I have already criticised where they are explicitly given us in advance (§52a-d above). There are other items introduced also which are not explicitly given us in advance.

The combined account, as its name implies, does not give us a single explanation of akrasia; not so much any explanation, in fact, as a can of worms. It contains intimations of at least 6 quite distinct ways of accounting for akrasia, some of which are not compatible with the others:

- (i) 'Biological' or 'compulsive' akrasia (a24, b4-9, and cp. 1151a20-24);
- (ii) 'Syllogism conflict' akrasia (and
- (iii) moral dilemma) (a25-36);
- (iv) 'Unconscious' akrasia (b11-13, and 1146b31-36);
- (v) 'Not real knowledge' akrasia (a36-b4, b9-11, b13-18);
- (vi) 'Syllogism-breakdown' akrasia (one hint in b11, and cp. 1146b36- 1147a10).

(i) The passage begins with words that we might translate, though I have never seen them so translated, as follows: 'Or again, one could view the cause of akrasia as being *biological*-in this way...'. This tendency to account for akrasia in a *biological*, or *physiological*, manner is quite marked in Aristotle:

'Fits of rage, and sexual desires, and certain other similar passions, quite evidently alter even the body, and in some cases actually cause kinds of madness. It is clear, therefore, that we should say that akrateis are in a similar state. The fact that they say the kind of things that come from knowledge proves nothing. Men in these states I have mentioned recite proofs and the words of Empedocles. New students chant together arguments that they do not yet understand. But knowledge [, to be truly understood,] has to grow into your very nature (συμφυῆναι)- and this takes time. We might say that akrateis' words are like those of actors (ὑποκρινομένωνς).' (NE 1147a14-24; cp. 1151a20-24, ἔστι δὲ τις..., quoted on p.6)

Reason decides what to do, but is swept away by some compulsive passion. Its direction remains right; but its movement in that direction is about as effectual as it would be for me, once embarked on an Intercity 125 which is thundering down the railway line to London, to start walking down the train back towards Edinburgh. In a case like this, human rationality is no more than a *veneer*. To talk of it merely obscures what is really going on. The claim to want to do the right is just talk, even if sincere talk. The real and effectual power in the agent is some ἐπιθυμία or θυμὸς, which drags the agent about like a slave. The idea is one that would have appealed strongly to the Euripidean Romantic Irrationalists Aristotle seems so against (NE 1110a28). If διάνοια δ' αὐτῆ οὐθὲν κινεῖ (1139a36), and it is desire (or passion) that κινεῖν γὰρ ἕκαστον δύναται τῶν μορίων (1147a36), then we can hardly say that on *this* account alone ἔοικεν ὁ ἐζήτει Σωκράτης συμβάλειν (1147b15).

Biological

417 If the cause of this kind of akrasia is biological, is in one's φύσις, then this kind is not full akrasia. It is a kind of partial akrasia which resembles θηριότης, for objectionable behaviour which results from φύσις is, by definition, θηριότης (NE 1148b15-1149a20; cp. the description of animal 'motivation' at dMA 701a29-36). The similarity to drunkenness and sleep of this kind of akrasia is noted by Aristotle (NE 1147b7-9). But this very similarity suggests more a pathological state in which normal rational decision is impossible, than anything like the clear and conscious decision to do wrong which is a constitutive part of full akrasia.

like The victim of such a pathological state is not clearly a voluntary agent at all, at the time when he is such a victim. Aristotle would say that it was up to the agent not to become this sort of person (NE 1114a4-7), or in the one-off case, not to get drunk on this occasion (1113b33-35), or otherwise to avoid situations of temptation. But if such an argument is meant to persuade us that such a case is a case *of* (full) akrasia, rather than one *resulting from* akrasia, it misses the point. The cases where the agent made the decision to get drunk, or to become the kind of person he now is, are clearly closer to full akrasia than anything the agent does while in the fit of passion or booze.

of. Action on such compulsions is plainly not full akrasia. In my schema, it will at most be either (1235) or (123).

moral

alter (ii) There seems no necessary connection between this biological argument presented at a24 and b4-9, and the very different argument about the form of the practical syllogism involved presented at a25-36. The 'two syllogisms' idea helps us see how close to being rational partial akrasia can be. But it is not obvious why this 'two syllogisms' shape, where one of the syllogisms is simply overridden by the other through the action of desire, should always or necessarily fit together with a biological compulsion. It is hard to imagine a case of akrasia where there were two conflicting syllogisms but no compulsion. Such a case would not really be a case of (even partial) akrasia. But it isn't hard to imagine a case where there was a compulsion, but no 'two syllogism' form. Such a form would obviously be less

like full akrasia, inasmuch as the akratic action is not rationally chosen. But my point is simply that the 'two syllogisms' explanation and the compulsion explanation are not necessarily parts of the *same* explanation.

When there are two syllogisms present to me, either I will recognise that I agree with the universal premiss of one but not the other, yet still end up acting on the one I disapprove of. This will be a very rational form of akrasia like (12345)- but it won't be full akrasia. Or

(iii) I will not be sure *which* universal premiss I approve of, and vacillate. While I vacillate, I am not in a state which is very like even partial akrasia. My position is, rather, one of moral dilemma (such as (3456) or (456)). I am confronted with two alternatives (or more? -Why not?), both of which may seem good to me, or (for that matter) bad, by different and perhaps incommensurable looking standards. If I resolve this dilemma, and act accordingly, I am no kind of akrates. If I resolve it, yet act against my resolution of it, then what I will be undergoing is likely to be (12345). If I act while this dilemma is still unresolved, then again it is unclear that I am any kind of akrates, for condition 1 is not satisfied: I am not clear that I disapprove of what I do, even if I am not clear that I approve of it either. Such a case will be (3456), (456) or the like.

(iv) 'Unconscious' akrasia (b11-13; cp.1146b31-36 and EE 1225b11-13). This, again, is not inconsistent with biological

akrasia; but it is not a necessary concomitant of it either. A physiological compulsion may cause me to lose sight of what I normally know to be right; but, as I said, it may also carry me protesting on. Getting myself in a state, like drunkenness, where I 'forget myself' is often combined in practice with getting into a state where I cannot control myself; but need not be.

As I have already argued, full akrasia could not involve anything like what Aristotle means by merely potential knowledge, or unconsciously held knowledge. While I am unconscious of an item of knowledge, I am, for all practical purposes, *ignorant* of it.

(v) 'Not real knowledge' akrasia (a36-b4, b9-11, b13-18). If the suggestion is indeed (as it appears to be) that there are two grades of knowledge possible, and that the type which gets 'dragged about like a slave' is only the lower grade, then it seems inconsistent of Aristotle even to mention the idea. He himself has presented plausible arguments to show the implausibility of the (Platonist?) suggestion that, in akrasia, it is only δόξα, and not ἐπιστήμη, which gets dragged around (1146b24-31). But at b16-18 he seems to be arguing something very like this himself:

'It does not seem to be the *true* sort of knowledge (τῆς κυρίως ἐπιστήμης) that is present when the fit (νόθος) [of akrasia] occurs; nor is it true knowledge that is dragged about by the fit, but perceptual (τῆς αἰσθητικῆς).'

This is hardly an argument at all; more like a furtive reversion to an opinion that Aristotle knows he is not really entitled to hold. For of course the Platonists' δόξα was precisely the class of forms of awareness like perceptual awareness; so if Aristotle is really espousing this opinion, he is siding with his own enemies.

Even if it were true that perceptual awareness were somehow second-rate as knowledge, it is hard to see how this could help explain akrasia. After all, as Aristotle himself remarks, 'Many who only have beliefs yet have no doubts, but think themselves to have exact knowledge' (1146b26-27). A belief held with perfect conviction, and never disproved, is practically speaking as good as knowledge; moreover it is the best we ever get in the area of perceptual knowledge; so it does not cut much ice to say that action against a syllogism based on such opinion is not *really* akrasia. Thus it is highly unconvincing to suggest (as Aristotle does at b16-18) that an action against the reports of perceptual awareness makes for a less unlikely form of akrasia than an action against the reports of knowledge of principle. For full akrasia would not primarily be action against the reports of *perceptual* awareness, but against those of *moral* awareness. In fuller akriasias than Aristotle gives us here, the point is nothing to do with perceptual awareness; it is that I do what I know is wrong. To think that these cases can be accounted for by a story about acting against perceptual awareness is to duck the real question.

Typically, 'not real knowledge' akrasia will be action in (effective) ignorance, and will therefore come no closer to full akrasia than (1234).

(vi) 'Syllogism-breakdown' akrasia (one hint in b11; cp. 1146b36- 1147a10). The hint I mean is:

'This opinion the akrates either *does not possess* or possesses it only so as to be able to repeat Empedocles like the drunkard does...'

Once again: to possess some knowledge, but only potentially (in the sense that Aristotle means), and not to possess it at all are not, where the explanation of action is concerned, distinct. Aristotle himself seems to blur the distinction here, in spite of the fact that the 'unconscious akrasia' argument depends on it.

If the akrates does not have the right particular premiss at all, not even potentially, then the syllogism breaks down. Either the akrates will have no opinions on the object of perceptual awareness in question, in which case (a) he is ignorant and (b) no syllogism, right or wrong, is possible; therefore his case is not one of full akrasia. Or else he will have a wrong opinion, in which case a wrong syllogism is possible. But since the wrongness of this syllogism derives from a particular piece of ignorance, this is not full akrasia either.

Thus 'syllogism-breakdown' akrasia is action in ignorance

too, and so no closer to full akrasia than (1234).

To sum up: Aristotle's combined account of akrasia is very confusing. This is largely because it is very confused. A variety of different ways of explaining akrasia are being employed simultaneously and promiscuously: so much so that one is often unsure which is being used. If the attempt is, as I assume, to show (irrelevantly to Aristotle's main concerns) that knowledge need not be supposed to be dragged around like a slave in the kind(s) of akrasia that really occur, then it must be judged a failure. If the attempt is (but it plainly isn't, unfortunately) to show how very profuse the sub-species of partial akrasia are, and how many ways we *might* explain some psychological phenomenon and still call it (partial) akrasia- then it succeeds. But, unfortunately for Aristotle, this is my thesis, not his.

3. The Varieties of Akrasia

So much for Aristotle's catalogue. Now for mine. As a coda to this essay, I will list some interesting forms of partial akrasia, and conditions like partial akrasia, drawn from my array of the sixty four permutations of C1-6. It will be obvious that none of these conditions (some of which, though not all, Aristotle mentions) counts as full akrasia in my sense, since none of them satisfies all of C1-6.

(Remember first that we know that some forms of akrasia, viz.

those involving the satisfaction of both C1-3 and C6, are logically impossible: which rules out not just (123456) but also (12346), (12356), and (1236).)

I. Failure of Condition 1

((23456), (2345), etc.)

1. Plain and simple ignorance that one is doing wrong, (a) when one isn't, (b) when one is.

Whether the others who think that I am doing wrong are right to think so or not, still I am plainly not acting against my own moral beliefs in this case. Ignorance is worth mentioning mainly because there are often cases where apparent akrasia can be explained by ignorance.

2. Incommensurability dilemmas.

If I genuinely accept the claims both of filial duty and of patriotism, I may face a dilemma when my choice is to fight with the Free French or care for my aged mother. I may truly *not know* what I should do, and hence form no unconditional practical imperative. But if (a) I resolve my dilemma, and act accordingly, then this is straightforwardly not akrasia. If (b) I resolve my dilemma by forming a UPI, but act contrary to my UPI, then this is a form of akrasia which does not satisfy C6. And if (c) I fail to resolve my dilemma and form no UPI, but still act on one or the other of my CPIs, then again my action cannot satisfy C6,

since I act on no UPI. (I do not act against my better judgement; I haven't got a (firmly held) better judgement.)

The subject of moral dilemmas and incommensurability is very interesting, but I doubt that that subject is central to the study of akrasia. Apart from the above considerations, the study of akrasia is a study of so many different phenomena that, in a way, *nothing* is central to it.

3. Uncertainty about a moral principle, which plays the part of an unconditional practical imperative, leading one to act against it. (Mere belief being mastered.)

Contra Aristotle (NE 1146b24 ff.), there seems no good reason why the difference between (mere) belief and (supposed) knowledge should not make a good deal of difference to how easily I slip into akrasia. Of course Aristotle is right that 'Some men hold no less firmly to what they only believe than others to what they actually know- as Heraclitus makes clear' (NE 1146b29). But this is just a statement of the difference between objective certainty (here OC) and subjective certainty (here SC); and there are some other interesting distinctions around.

SC is *feeling* sure that p, the certainty found in some kinds of religious faith, and not necessarily supported by anything outside itself. OC is *having good reasons to be* sure that p, the kind that was supposed by Logical Positivists to be most characteristic of the certainties of mathematicians and

scientists. OC and SC are not mutually exclusive. OC often engenders a sense of SC; SC makes us look for reasons, which if we find them may lead to OC.

Aristotle argues that we can't say that the difference between SC alone and SC + OC is significant for akrasia, because akrasia is a matter of what goes on in the agent's own deliberations. Therefore what matters for akrasia is SC. This, no doubt, is correct; but what about the possible *degrees* of SC? I am (let us say) strongly tempted to go to bed with my neighbour's wife. But I also have a firm conviction, a matter of complete SC, that it would be wrong to do so. In these conditions, it is (*prima facie*) most unlikely that I *will* go to bed with my neighbour's wife. But suppose that my conviction that I should not bed my neighbour's wife is weaker than complete SC. Surely it is plausible to say that, the weaker my anti-adultery conviction, the more likely I am to end up an adulterer? So, in the sense of 'believe' that we oppose to 'know' ('You don't *know* that, you just *believe* it'), it *does* seem that a distinction between moral knowledge and moral belief is relevant to some forms of partial akrasia. So, for example, with Pierre Bezuhov in *War and Peace*:

"It would be nice to go to Kuragin's," he thought, but immediately recalled his promise to Prince Andrei not to go there again. Then, as happens to people who have no strength of character, such a passionate desire came over him for one last taste of the familiar dissipation that he decided to go. And the thought immediately occurred to him that his word to Prince

Andrei was not binding because before he had given it he had already promised Prince Anatole to come. "Besides," he reasoned, "all these 'words of honour' are mere convention and have no precise significance, especially if one considers that by tomorrow one may be dead, or some extraordinary accident may happen to sweep away all distinctions between honour and dishonour." Arguments of this kind often occurred to Pierre, nullifying all his intentions and resolutions. He went to Kuragin's.'

(Tolstoy, *War and Peace*, Ch.1)

We may note, incidentally, that there are many ways in which my subjective certainties about right and wrong can be eroded. One of them is if I am given conflicting injunctions by different moral authorities. Another is if I witness hypocrisy in those who seem most strongly to hold the moral principles which I also hold.

4. Failing to see that *this* action is covered by a moral principle of which one is aware of.

I may hold a UPI that actions of type P are to be avoided; but I fail to see that action p is of type P. Hence my UPI is not 'transferred' to the conclusion of my practical syllogism. But this is just a kind of ignorance of a particular premiss necessary to a correct syllogism of the type favoured by Aristotle.

5. The *suppression* of one's awareness of the truth of the belief expressed in Condition 1.

This is the first reference I have made directly to another interesting phenomenon which deserves, and receives (but not here), massive treatment of its own: the phenomenon of *self-deception*. The phenomenon was evidently known to St. Paul (οὐκ ἔδοκίμασαν τὸν θεὸν ἔχειν ἐν ἐπιγνώσει... , Romans 2.28). Perhaps, even, it is the foolishness of what he takes to be the self-deception involved in the rejection of an instinctive awareness of God that made the Psalmist say that it is the *fool* who 'hath said in his heart, There is no God' (Psalm 53.1).

In more recent centuries the idea of self-deception has been of importance to thinkers as important as Butler, Marx, Freud, and Sartre. As Patrick Gardiner is not the first to have pointed out, the idea is paradoxical:

'It is surely odd to suggest that somebody could try to make, and succeed in making, himself believe something which he, *ex hypothesi*, at the same time believes not to be true.'

(Gardiner 1969/70)

The logical perplexities of self-deception seem remarkably similar to those involved in some kinds of partial *akrasia*. Perhaps we could call it *doxastic akrasia*, when someone deliberately sets about giving himself new and less inconvenient

beliefs: persuading himself that adultery isn't really wrong, or alternatively that this isn't really a case of adultery, etc. (see cases 36a-i).

But if we say that I do this, fully consciously, to myself, and am successful and hence become an akrates in my action, an obvious rejoinder follows: 'No, the action was not a case of (full) akrasia, but simply done under a delusion. What was (full, or fuller) akrasia was the act of persuading yourself to hold false beliefs.' (Gardiner: 'One way out... consists in arguing that self-deception is really a form of other-deception... On such a hypothesis, I am prevented (I do not prevent myself) from recognising certain things about myself... [But the] difficulty merely re-emerges... If [the censor] is to perform the functions assigned to it... it must presumably be aware of its own activity... But this implies that the censor in its turn is in "bad faith".')

When I consciously make myself unconscious of certain facts of which I am conscious but wish to be unconscious, is this full akrasia? If I choose to suppress facts, then presumably I formulate a UPI to do so, and do so. But this does not necessarily involve me in acting against a UPI. If I do act against a UPI in self-deception, then clearly my act of self-deception cannot satisfy C6. If I do not, then it cannot satisfy C1. Either way, self-deception cannot involve full akrasia.

Doxastic akrasia, I suggest, is an important form of partial akrasia; but (as with moral dilemmas) the temptation should be resisted to make this account the model for all forms of explanation of akrasia. For there are plenty of cases where my akrasia has nothing to do with my suppressing my own beliefs. 'Partial akrasia' can mean being dragged unwillingly about by passion, or denying with complete sincerity the truth of some crucial premiss of the syllogism, just as well as it can mean self-deception.

6. Problems of various sorts about the distinction between actual and potential knowledge (as discussed in §2).

II. Failure of Condition 2

(156) (?)

7. Behaviour ($\theta\eta\rho\iota\acute{o}\tau\eta\varsigma$) with no actual or potential conditional motivation towards the action which expresses akrasia.

This, surely, will only occur when one is behaving completely compulsively. Examples of completely compulsive behaviour are examples of full $\theta\eta\rho\iota\acute{o}\tau\eta\varsigma$; but they are also examples of partial akrasia.

7*. Not: 'Deliberate *suppression* of one's awareness of the truth of the belief expressed in Condition 2'.

To do X while actually suppressing one's belief that 'conditionally (in one way or another), X is to be done by A at t1', would be a very strange psychological phenomenon; but it would not count as akrasia.

III. Failure of Condition 3

(12456 and others)

8. Making an exception.

Cases like that of the person who can't resist the cream cake 'just this once' are often cases where a belief about the relationship between unconditional and conditional motivation is suspended or claimed not to apply on every occasion. One may make an exception (a) of this *occasion*; or (b) of oneself or a partner in crime, of this *agent*; or (c) of the dirty deed, of this *action*.

9. Amoralism.

Wherein one simply denies in general, without trying to make excuses, that the UPI of C1 does override the CPI of C2; and, normally, rejects the UPI. Just as type 7, on the one hand, is very close to full θηριότης, so this kind, on the other, is very close to full κακία.

10. Deliberate *suppression* of one's awareness of the truth of the belief expressed in C3.

V. my remarks on type 5.

IV. Failure of Condition 4

11. External duress.

12. External compulsion.

13. Internal duress.

14. Internal compulsion.

The difference between duress and compulsion was explained in Ch.1, §2d-f.

15-18: *Pretending* to be under one of these, or (19-31) more than one of them.

If the pretence (a) only fools *others*, the pretence itself is a kind of κακία. If (b) it even fools *me*, the remarks about self-deception for case 5 apply.

V. Failure of Condition 5

32a-i. Total ignorance of different parts of the practical syllogism.

This, of course, is never anything like full akrasia.

33a-i. Effective ignorance of different parts of the practical syllogism due to the relevant piece(s) of knowledge being only 'potential' (in Aristotle's sense).

34a-i. Effective ignorance of different parts of the practical syllogism due to the relevant piece(s) of knowledge being only 'unconscious' (in Aristotle's sense, whatever that is).

35a-i. Effective ignorance of different parts of the practical syllogism due to *deliberate suppression* of the relevant piece(s) of knowledge.

V. case 5.

36. Failure to formulate a practical syllogism correctly because of a fault in the reasoning: i.e., by drawing a wrong inference from good premisses by committing any one of the kinds of fallacy that apply to practical reason (for these, see Ch.3, 54).

37. Formulating a practical enthymeme instead of a practical syllogism.

This I find an interesting possibility. If rhetorical arguments aim (subjectively) at persuading us rather than

(objectively) at establishing the truth (dAR 1356a1-4), isn't it possible that what happens, quite often, in akrasia is that I listen to and act on an emotionally appealing argument because it is easier to do so than to listen to a logical argument? Of course, if I do so, I am not acting fully rationally, and so not fully akratically.

VI. Failure of Condition 6

38. Irrationality is due to only one cause: the lack of an explanation which applies to one's behaviour. See above, §1.

And that's the end of my catalogue. Of course, some of these categories overlap; but on the other hand, several of them are subdivided into different alternatives. I started by showing that, in a defined sense, there *is* no *full* akrasia, and hence no great threat to Aristotle's theory of proairesis. This last catalogue, my catalogue, is my final demonstration that *partial* akrasia is a far more complicated affair than most philosophers give it credit for- including, but *not* especially, Aristotle. My original suggestion was that the subspecies of akrasia were to be found among a class of up to 2^6 types. That is an idea I have not deserted. But this last catalogue suggests strongly that, for each condition except C6, there is not just one way in which it might fail to be satisfied.

If we bear in mind that not many of the 38 states of partial

akrasia described in my catalogue are mutually incompatible, and that many of these states are subdivided into different sub-possibilities, we will see that most of these states could occur *together*. Which will give sense to my final suggestion. The subspecies of akrasia were to be found among a class of up to 2^6 types; but evidently the *sub-subspecies* of akrasia are to be found among a class with something of the order of 2^{38} members. If we accept Aristotle's opinion that the akrates is, at least sometimes, at least ἡμιπόνηρος (1152a18), we will then be in an even better position than Aristotle himself was to appreciate how very applicable to akrasia is the untraced line of verse which he quotes so approvingly at NE 1106b35:

ἔσθλοὶ μὲν γὰρ ἀπλῶς, παντοδαπῶς δὲ κακοί.

Table Two: The 64 permutations of the Conditions of

Akrasia

0	1	2	3	4	5	6
		12	23	34	45	56
			13	24	35	46
				14	25	36
					15	26
						16
			123	134	145	156
				124	135	146
					125	136
						126
				234	245	256
					235	246
						236
					345	356
						346
						456
				1234	1245	1256
					1235	1246
						1236
					1345	1356
						1346
						1456
					2345	2356
						2346
						2456
						3456
					12345	12356
						12346
						12456
						13456
						23456
						123456

Chapter Five:

Freedom and the Voluntary in Augustine's 'de Libero Arbitrio'

1. Introduction
2. Book One
3. Book Two
4. Book Three

'The concept of 'free will'... is the most infamous of all the arts of the theologian for making mankind 'accountable' in his sense of the word, that is to say for *making mankind dependent on him*... The doctrine of will has been invented essentially for the purpose of punishment.'

(Nietzsche, *Twilight of the Idols*, tr. Hollingdale, p.53)

'If it is not by the exercise of will that we do wrong, no one at all is to be censured or warned. If you take away censure and warning the Christian law and the whole discipline of religion is necessarily abolished. Therefore it is by the will that sin is committed.'

(Augustine, *de Vera Religione* 27, tr. in Burleigh, 1953)

1. Introduction

The idea developed in my study of Aristotle has been that freedom to act means no more and no less than voluntary action, which in turn is defined as action which satisfies conditions about compulsion, ignorance and what I have called 'irrationality' or 'rational inexplicability'. I have also offered, if not a resolution, then at least a classification, of the various kinds of problems there are about akrasia. I have suggested that some of these problems are real, but others are not. In particular, there is no problem about answering the question 'How is it possible for someone to do fully voluntarily what she honestly and seriously believes it better not to do?'; because, in Aristotle's terms, this isn't possible.

It will be interesting to attempt to apply this approach to Augustine's theory of freedom and the voluntary. Will this approach work for Augustine? Will the project yield consequences for Augustine's notion of deliberate wrongdoing which are anything like analogous to the consequences I drew about Aristotle's notion of akrasia?

Augustine, like Aristotle, offers us some remarks about freedom and the voluntary which are not readily separable from the mosaic of other issues with which he is usually concerned. Unlike Aristotle, Augustine also offers us a set-piece exposition, the *de Libero Arbitrio*, which purports to be exclusively concerned with the theme of freedom of action. It

would be an ungrateful exegete who rejected this invitation to begin his exploration of the theme of human freedom in Augustine where I shall begin it: with a paratactic survey of the *de Libero Arbitrio*.

The question 'Unde malum?' (Conf 3.7) is a standard problem for Jewish and Christian apologetics; and there is a standard answer to it. To quote the Epistle of James, 1.13-15:

μηδείς πειραζόμενος λεγέτω ὅτι Ἐκ τοῦ θεοῦ πειράζομαι· ὁ γὰρ θεὸς ἀπειραστός ἐστίν· κακῶν, πειράζει δὲ αὐτὸς οὐδένα. ἕκαστος δὲ πειράζεται ὑπὸ τῆς ἰδίας ἐπιθυμίας ἐξελλκόμενος καὶ δελεαζόμενος· εἴτα ἢ ἐπιθυμία συλλαβοῦσα τίκτει ἁμαρτίαν, ἢ δὲ ἁμαρτία ἀποτελεσθεῖσα ἀποκίει θάνατον.

'The free will defence' is a gambit in Judaeo-Christian theodicy as old (at least) as the third chapter of *Genesis*. The philosophical interest of Augustine's presentation of it in the *de Libero Arbitrio* lies not so much in the recapitulation of this old, old theme, as in the theoretical underpinnings which Augustine gives it; especially, the philosophy of action and intention in the context of which he interprets it. At least in the degree to which Augustine organised and systematised this philosophical context, Augustine was a radical innovator in the Judaeo-Christian tradition of 'free will' theodicy.

The dLA (387-388 A.D.) is written in the form of a philosophical dialogue, or, better, set of three dialogues (v.

1.35 'aliud tempus sumendum est', 2.1 'in superiori libro...', 2.54 'in aliam disputationem differendum est', 3.1 'si opportunum existimas...', 3.2 'superioribus duabus disputationibus'). These dialogues between Augustine and his friend Evodius present, against Manichaean teaching, expositions on three important issues for 'free will' theodicy:

Book One: Why is there any evil in the world at all?

Book Two: If the world's evils result from the gift of free will, should God have given humanity this gift?

Book Three: If God should have given this gift, how did it come about that humanity misused it?

It is in the context of these inquiries that I will pursue the question which concerns me: What *is* free will, for Augustine?

The division given above is only a rough guide to the complexities of the dLA. Only at Ch.24 (out of 35) does the first book begin to address directly the proof of James's thesis that 'Quisque malus sui malefacti auctor est' (dLA 1.1). Over half of the second book (2.7-39) is devoted to proving God's existence, not (directly) to justifying the gift of free will. The third book discusses foreknowledge and predestination (3.4-10) and the nature of the universe (3.12-26) as well as the fall of humanity (3.47-77). But I begin with Book One.

2. Book One

Summary of Contents:

1.1-3: Moral and natural evil; natural evil disposed of. On moral evil: first suggestion: is it taught? -No.

1.4-5: Methodological remarks: 'Credo ut intelligam'.

1.6-15: Definition of sin: based on law. To be in accordance with eternal law is to be in perfect order.

1.16-18: Definition of 'order': the threefold distinction (esse, vivere, intelligere); the hierarchy of nature.

1.19-21: Wisdom, reason and passion.

1.22-24: Sin its own penalty; but is this just?

1.25-30: The doctrine of the supremacy of the good will; its relation to the cardinal virtues.

1.31-34: Will and law; 1.34 the two genera of men.

1.35: Restatement of the definition of sin; *avant-propos* of Book Two.

1.1-3: Moral and natural evil. At the beginning of the dLA Evodius puts to Augustine the question which the *Confessions* tells us he wrestled with: 'Dic mihi, quaeso te, utrum Deus non sit auctor mali?' Augustine's first response is to make a distinction:

'Duobus enim modis appellare solemus malum: uno, cum male quemque fecisse dicimus; alio, cum mali aliquid esse perpessum.'

(1.1)

The evils we suffer get exceedingly short shrift in the dLA. God, apparently, is the author of this kind. From the premiss that 'Divina providentia hoc universum regi', Augustine thinks it follows that 'Nemo iniuste poenas luit', and hence, since obviously there is suffering in the world, that such suffering must always be punishment for *something*. How long Augustine would have stuck by this outrageous argument is not made clear. The underlying idea, that the universe is in some sense perfect exactly as it is, which Augustine returns to at dLA 3.24, seems more akin to Stoic thought than to Christian. Compare Long (1974), p.165:

'The Stoics held that this is the best of all possible worlds; notwithstanding apparent imperfections here and there, Nature so organises each part that harmony is present in the whole. The psychological and moral implications of this notion are constantly invoked by Marcus Aurelius, and it seems to be a fact that many men have found comfort in the belief that, come what may, their lives contribute to some grand universal scheme: "Everything that is in tune with you, O Universe, is in tune with me. Nothing that is timely for you is too early or too late for me" (Aurelius, *Meditations* 4.23).'

But what of the other kind of evil?

'*Evodius*. Est ergo auctor illius mali, cuius Deum non esse compertum est?

Augustinus. Est certe: non enim nullo auctore fieri posset... non

[autem] unus aliquis est, sed quisque malus sui malefacti auctor est.' (dLA 1.1)

As I have said, this is an uncontroversial claim, at least to orthodox Christians. The interesting question is how to establish it.

Augustine begins this task by suggesting that evil cannot be *learnt* (1.3). If it were, of course, the responsibility for my evil would simply revert to my teacher in evil, and we would have to ask whether *he* learnt evil, and if so from whom (a regress which might be suggested by the role of the serpent in the garden of Eden). Augustine has other arguments to make his point, however: principally the intellectualist claim that, since learning is necessarily of the good, evil is exactly that of which there cannot be learning.

(As Burnyeat (1987) has pointed out, Augustine argues in the *de Magistro* that, in one sense, no one else can ever teach me anything, good or bad; I have to 'cotton on' for myself, which is essentially an internal business, like *anamnesis* in the *Meno*. This need not contradict Augustine's argument here: he can still say that this kind of 'cottoning on' only happens with learning of the good.)

1.6-15: *Sin, Law and Order*. If not by teaching, then how do we come to do evil? Augustine's eventual answer to this will be that '[male] facimus ex libero voluntatis arbitrio' (dLA 1.35).

He does not set his course towards this claim very directly, though, of course, one argument for it has been given already (1.1):

'Malefacta iustitia Dei vindicari. Non enim iuste vindicarentur, nisi fierent voluntate.'

But this begs the question of *which* deeds, exactly, are the evil ones punished by God's justice. Perhaps this is why Augustine moves on, at 1.6, to the definition of 'malefacta'.

Evodius makes two suggestions about what the essence of evil-doing could be: (i) an appeal to the 'do as you would be done by' principle (1.6); (ii) an appeal to what society generally condemns (1.7). Augustine rejects both, and also his own suggestion that 'Fortassis ergo libido in adulterio malum est' (1.8). Some desires ('libidines' or 'cupiditates') are shared by all humans. But there is a vital difference between the desires of good and bad humans:

'Cupere namque sine metu vivere, non tantum bonorum, sed etiam malorum omnium est: verum hoc interest, quod id boni appetunt avertendo amorem ab iis rebus, quae sine amittendi periculo nequeunt haberi; mali autem ut his fruendis cum securitate incubent, remove impedimenta conantur, et propterea facinorosam sceleratamque vitam, quae mors melius vocatur, gerunt.' (1.10)

Augustine's principle- and it is an odd one: his is, as Nussbaum (v. Nussbaum 1986) might say, not a very fragile good- is that *all* desire for *anything* that I could conceivably lose against my will, is wicked.

'Culpabilis cupiditas... esse iam apparet earum rerum amor, quas potest quisque invitus amittere.' (1.10)

As Evodius points out (1.13), this definition of 'malefacta' conflicts with the earlier requirement that the authority of law should be upheld by Augustine's definition of 'malefacta' (1.6). For the laws enforce us in possession of impermanent goods. Augustine's response (1.13-14) is to introduce a distinction between the human law and the divine- an early hint of his later 'two cities' theory. The good person is the one who lives in accordance with the divine, not the human law. But what does this mean?

1.16-23: *The threefold distinction.* Accord with the divine law turns out to consist- shades of the Stoics again- in Order: 'iustum esse omnia ordinatissima' (1.15). And what is Order? Augustine proposes (1.16-19) a philosophy of mind based on the threefold division vivere/ sentire/ intelligere. Humans are rightly ordered when reason ('intelligentia') is in control, not 'sensus' or mere vegetable life.

Now we begin to move back towards the discussion of freedom. For when a human is rightly ordered, his mind controls his

physical nature. Then nothing external to him, either physical or spiritual, can compel him to do evil. Nothing except God is more powerful than a rightly ordered mind (1.20), and God is good. Vicious minds are weaker, and therefore cannot compel the rightly ordered mind. Other rightly ordered minds, by definition, will not want to lead a rightly ordered mind into disorder:

'Quidquid par aut praelatum est [regnanti menti compotique virtutis] non eam facit servam libidinis propter iustitiam... nulla res alia mentem cupiditatis comitem faciat quam propria voluntas et liberum arbitrium.' (1.21)

And yet, points out Augustine, 'iuste illam [mentem] pro peccato tanto poenas pendere' (1.21). Human minds (or beings) are punished for doing wrong; therefore, by the argument of 1.1 again, it follows that this punishment must be just; and, by the argument of 1.16-21, it also follows that the fault of an unfallen being like Eve *cannot* have originated outside her. So it must have originated inside her...

1.22-30: *The supremacy of the good will*. At this crucial point Evodius interrupts with two importunate questions: (i) How could someone *choose* to descend from the 'sublimitate sapientiae'? (ii) Have we ever been wise enough to avoid falling into evil? (ii) raises the issue of the origin of the soul (1.24), and suggests a paradox which Augustine eventually tackles in 3.73. It is Augustine's concentration here on (i) which introduces the first explicit discussion of the will (1.25-30).

'Sitne aliqua nobis voluntas?', asks Augustine; Evodius' answer is 'Yes' (1.25). All humans, it is suggested, have 'voluntas'. Moreover, all humans have '*bona voluntas*', and indeed the possession of 'voluntas' apparently implies the possession of at least some sort of '*bona voluntas*'. For all humans desire-
'modo tu vide!'- '*recte honesteque vivere, et ad summam sapientiam pervenire*' (1.25)- however confusedly they may desire it. This good will is the only thing which (in Kant's words, the similarity of which is noted by Thonnard (1941), p.496, and Burnaby (1938), p.183) 'can be called "good" without condition':

'Vides igitur iam, ut existimo, in voluntate nostra esse constitutum, ut hoc vel fruamur vel caremus tanto et tam vero bono.'

And hence Augustine urges us to treat the good will itself as the object supremely worthy of pursuit:

'Quid enim tam in voluntate, quam ipsa voluntas sita est? Quam quisque cum habet bonum, id certe habet quod terrenis omnibus regnis, voluptatibus omnibus corporis longe anteponendum sit. Quisquis autem non habet, caret profecto illa re, quam praestantiorem omnibus bonis in potestate nostra non constitutis, sola illi voluntas per seipsam daret. Itaque cum se ipse miserrimum iudicet, si amiserit [bona mundi], tu eum non iudicabis, cum iis inhaeret quae amittere facillime potest, neque dum vult habet, caret autem bona voluntate, quae nec comparanda est cum istis, et cum sit tam magnum bonum, velle solum opus est,

ut habeatur?' (1.26; my italics)

It might be said that there are two reasons for the supreme value of the good will as an object of desire. First, since 'bona voluntas' is defined as that voluntas 'qua appetimus recte honesteque vivere, et ad summam sapientiam pervenire' (1.25), it follows that the rest of the Good Life will be impossible without it. And second, we might add, since to *want* to have a good will is *eo ipso* to *have* a good will, it is something that can be had simply by desiring it, and hence it cannot be lost against one's will. This makes it, consistently with the argument of 1.10, preeminently the *right* object of desire.

These two points are both suggested by Augustine's remarks in dLA 1 and sometimes elsewhere. But to some extent they are in competition. The first point is that the good will's activity is only necessary for the good life; the second is that it is both necessary and sufficient. The second point may be coordinated with a general view of 'voluntas' as being undetermined by anything outside itself; the first with a general view which sees the activity of any 'voluntas' as inescapably cohering with a much larger scheme of things. I shall, in the end, reject the second point and its concomitant view, and argue for the first. But let me first explore what I mean by the second.

'Voluntas', for Augustine, is (it might be said) something that can not only act on other extensional objects (as means) and towards other intentional objects (as ends); 'voluntas' can also

act on and towards itself. The will at time t1 can take, possibly its own present state, but more likely its own future state at some t2, as an intentional object of desire. For example, at 1.26, the will is seen as having the capacity to will that it should become a good will, which is unmistakably an aspiration to a future intentional object.

'...Quam [rem] [sc. bonam voluntatem] sola voluntas *per seipsam* daret.' (1.26)

'Nonne bonam voluntatem suam diligere, et tam magni aestimare quam dictum est, *etiam ipsa bona voluntas est?*' (1.28)

This intentional object of the will is some state of affairs which includes the goodness of the same will. This willing is reflexive or *self*-referential in the sense that it is (we are to presume) one and the same self which does the willing at t1, and thereby attains the desired state at t2. In this act of willing the will in question is both subject and object.

'Hanc igitur voluntatem, si bona itidem voluntate diligamus atque amplectamur, rebusque omnibus quas retinere non quia volumus possumus, anteponamus; consequenter illae virtutes, ut ratio docuit [v. 1.27], animum nostrum incolent, quas habere idipsum est recte honesteque vivere.' (1.29)

From such words as these it is natural, on the second view, to draw out the claim noted above, that in the very act of

willing to have a good will and by virtue of my so willing, I immediately and by definition have what I will. For we remember again that the good will is defined as that 'qua appetimus recte honesteque vivere, et ad summam sapientiam pervenire' (1.25). Now this definition will clearly include (i) any act of the will which has a good (first order) object. Hence, since the (second order) act of the will which is willing to have a good will is itself an act of the will which has a good object, (ii) as soon as I make this second order act of the will, I have a good will.

It may be observed that this kind of analysis can be applied equally to the *bad* will (indeed, Augustine suggests such an application at 3.48 ff.). At dVR 26 we read that 'the bad angel loved himself more than God, refused to be subject to God, swelled with pride, came short of supreme being, and fell'. In that order? Is this a series of causes and consequences: or is its end in its beginning? If it is a chronological sequence, at which point in this sequence did the bad angel fall? Surely at the very beginning of it. On any orthodox view, one can hardly love oneself more than God without already having committed the worst sin there is. It seems inappropriate to put '...and fell' at the very end of this story of already pernicious acts. In fact, it looks wrong to call this a chronological sequence at all. It makes more sense to see Augustine's 'and's here as implying a number of different effects which came about simultaneously, being not causally but logically involved with each other.

This would fit well with the notion of the reflexivity of the will which is currently being canvassed. For what we may say about the bad angel on that view is that, as soon as he willed to have a bad will, he had one, instantaneously and exactly in virtue of so willing. This bad will need have no antecedent in the way that each element in the above sequence, including the first, seems to require. Hence (dLA 3.48-49) it might seem that there is no need to postulate anything that precedes the act of the will as a necessary and sufficient condition, and hence there is no problem about God's alleged responsibility, direct or indirect, for the iniquities of free agents. If Augustine can make this notion of the 'reflexivity of the will' work, he will have provided an elegant solution to his problem of theodicy, by making the will *its own cause*, a self subsistent entity.

An interesting parallel to this idea of reflexivity in Augustine's philosophy of mind is that one of the grounding certainties, for the Augustinian person, is her self-conscious awareness of her own mind (1.16): hence Augustine's well known anticipation of Descartes, 'Si fallor, sum'. When we add to this the point that knowledge too 'knows itself' we have the trinity of introspection of *de Trinitate*, the *Confessions*, and the *City of God*:

'Augustine insists that introspection yields an irreducible trinity. I am, I know, I will: I am a being that knows and wills; I know and will my being, my knowledge, and my will.'

(Burnaby (1938), p.144-5)

Now we might wonder: Is there such a thing as an 'act of the will'? Do we *do* something when we will, as an exordium to doing anything else? Or should we take a non-causal view of the relation between will and action, and see talk about the voluntary as being rather talk about the way *one* thing is done (viz., a voluntary action) than about whether a certain relationship holds or fails to hold between *two* things that are done, viz., the willing (or lack of it) and the action in which it issues? If Augustine's view is that acts of the will are causes of actions, the classic, Rylean accusation is that such a view generates a vicious regress:

'[Both] mental [and physical] processes can then, according to the theory, issue from volitions. So what of volitions themselves? Are they voluntary or involuntary acts of mind? Clearly either answer leads to absurdities. If I cannot help willing to pull the trigger, it would be absurd to describe my pulling it as "voluntary". But if my volition to pull the trigger is voluntary, in the sense assumed by the theory, then it must issue from another volition and that from another *ad infinitum*... The doctrine of volitions is a causal hypothesis, adopted because it was wrongly supposed that the question, "What makes a bodily movement voluntary?" was a causal question.'

(Ryle 1949, Ch.3)

However, the upholder of this understanding of Augustine need not be too disturbed by Ryle's attack. He might simply respond

that there is indeed a regress, but not a vicious one, in Augustine's analysis of the 'voluntas'; and perhaps he might connect the self-referential nature of the 'voluntas' with classical philosophy's notion of the 'self mover'.

A more telling criticism of this interpretation of Augustine's idea of the reflexivity of the will might be to question the conclusion of the interpretation, that we must see Augustine as having thought that all motions of the will are altogether undetermined by anything outside the will. This consequence alone gives us strong reason for rejecting as an adequate overall understanding of Augustine any interpretation which leads to it. As we shall see in Chs.8-9, Augustine is in fact crucially committed to *rejecting* this kind of indeterminacy of the will, at least as a general claim about 'voluntas'. For him the only indeterminate will is a bad will- as more careful exegesis will make plain. (V. my remarks on dLA 3.47 ff., below.)

Thus, whatever else Augustine may have meant by aligning 'velle' with 'esse' and 'intelligere' as one of a triad of mental functions which display some sort of reflexivity, it cannot have been the elegant doctrine just outlined, that the will is normally altogether undetermined by anything outside itself. Seen in its proper context, Augustine's notion of reflexivity is more likely to be put to a less ambitious use, as an attempt to articulate philosophically our ability to be related to ourselves as both objects and subjects of thought, being and choice. Nothing need follow, from this doctrine of reflexivity, about the

indeterminacy of choice, nor indeed of thought or being. In the terms of my analysis above, we can admit the claim that 'voluntas can act on and towards itself', without having to make the crucial transition to the further claim that 'in the very act of willing to have a good will and by virtue of my so willing, I immediately and by definition have what I will'. And reflection does suggest that the former claim is more plausible than the latter, which would seem to commit us to the idea that good will can be formed in a person instantaneously. Augustine's most thought-out beliefs about moral development, as we shall see, are much less crude than that. Such an idea suggests a 'perfectibilism' which is vigorously rejected by Augustine, e.g. at dCE 4.20, where Augustine warns us that 'anyone who thinks that in this mortal life a man may so disperse the mists of bodily and carnal imaginings as to possess the unclouded light of changeless Truth, and cleave to it with the unswerving constancy of a spirit utterly set apart from the common ways of this life—he understands neither what he seeks, nor what he himself is who seeks it'. (Compare Justin Martyr's rueful comment, ὀλίγου ἐντὸς χρόνου ᾧμην σοφὸς γεγονέναι, καὶ ὑπὸ βλακείας ἤλπίζον αὐτίκα κατόψεσθαι τὸν θεόν: Justin, *Dialogus contra Tryphonium*, 2.)

DAUER 11/14

1.27: *The good will and the cardinal virtues.* Augustine goes on to show the centrality of the good will to the good life in a different frame of reference, by cementing the doctrine of the will, the central icon in his (from a classical point of view) highly revisionary philosophy of mind, into place in a reputable classical alcove, the doctrine of the four cardinal virtues:

1. Virtue: Prudentia.

Definition: 'Appetendarum et vitandarum scientia'.

Relation to 'good will': 'Quisquis ergo bonam habens voluntatem... hanc unam dilectione amplexetur, qua interim nihil melius habet, hac sese oblectet, hac denique perfruatur et gaudeat... quamque invito illi eripi vel surripi nequeat... nullane hunc putamus praeditum esse prudentia, qui hoc bonum appetendum, et vitanda ea quae huic inimica sunt videt?'

2. Virtue: Fortitudo.

Definition: 'Illa animae affectio, qua omnia incommoda et damna rerum non in nostra potestate contemnimus'.

Relation to 'good will': '[Ille qui bonam habet voluntatem,] illa quippe omnia quae non in potestate nostra non sunt, amare iste ac plurimi aestimare non potest. Mala enim voluntate amantur, cui tanquam inimicae carissimo suo bono resistat necesse est. Cum autem non amat haec, non dolet amissa, et omnino contemnit; quod opus est fortitudinis.'

3. Virtue: Temperantia.

Definition: 'Affectio coercens et cohibens ab iis rebus quae turpiter appetuntur'.

Relation to 'good will': 'Utrum ab eo temperantiam alienare possimus, cum ea sit virtus quae libidines cohibet? Quid autem tam inimicum bonae voluntati est quam libido? Ex quo profecto intelligis istum bonae voluntatis suae amatorem resistere omni modo atque adversari libidinibus, et ideo iure temperantem vocari.'

4. Virtue: Iustitia.

Definitio: 'Virtu[s] qua sui cuique tribuuntur'.

Relation to 'good will': 'Qui enim habet et diligit voluntatem bonam et obstitit eis... quae huic iniuria sunt, male cuiquam velle non potest. Sequitur ergo ut nemini faciat iniuriam: quod nullo pacto potest, nisi qui sua cuique tribuerit'.

The life of the man of good will is the happy life, the 'good life' in Aristotle's sense. For the life of the man of good will is 'laudabilis'; and if it is 'laudabilis', then it is 'non fugienda' but 'appetenda sedulo'. Hence, says Augustine, since the praiseworthy life is to be sought after (one is reminded of the famous utilitarian equivocation over the word 'desirable'), it must be the happy life. And hence the happy life is the life of the man of good will. Conversely, this must mean that, 'etiamsi nunquam antea sapientes fuimus, voluntate nos tamen laudabilem et beatam vitam, voluntate turpem ac miseram mereri ac degere'.

Augustine has aimed to demonstrate that the life of the man of good will is the virtuous life (so it is up to us to be good), the praiseworthy life (so it is up to us to pursue it), and the good life (so it is up to us to be happy). He now wants to show that this life is also perfectly easy:

'...Conficitur ut quisquis recte honesteque vult vivere, si id se velle prae fugacibus bonis velit, assequatur tantam rem tanta facilitate, ut nihil aliud ei quam ipsum velle sit habere quod voluit.' (1.29)

(‘...si id se velle... velit’: note the reflexivity again.)

Of course, if this life is perfectly easy, that does prompt the question of why people *seem* to find it so difficult. What then is the difference between those who succeed in living the good life, and those who don’t? Augustine’s reply is that it is possible to seek either happiness or righteousness. But happiness is not what we should be looking for, but righteousness, of which happiness is as it were a ‘by-product’ (Mt.6.33).

‘Nam illi qui beati sunt, quos etiam bonos esse oportet, non propterea sunt beati, quia beate vivere voluerunt; nam hoc volunt etiam mali: sed quia recte, quod mali nolunt... Itaque cum dicimus voluntate homines esse miseros, non ideo dicimus, quod miseri esse velint, sed quod in ea voluntate sunt, quam etiam eis invitis miseria sequatur necesse est.’ (1.30)

1.31-34: *Will and law again.* At this point Augustine reintroduces the legal conception of morality:

‘Hoc enim aeterna lex illa... incommutabili stabilitate firmavit, ut in voluntate meritum sit; in beatitate autem et miseria praemium atque supplicium.’ (1.30)

This being so, which of the two, the good man and the bad, will love the ‘aeterna lex’? Augustine apparently argues in 1.31 that the good man will love it and the wicked hate it precisely because it rewards the one and punishes the other. (This seems to

subvert the very idea of absolute morality that Augustine is arguing for; he has just said that the good man is happy because he seeks to be good, not good because he seeks to be happy (1.30.)

In any case, Augustine thinks it now 'manifestum... alios esse homines amatores rerum aeternarum, alios temporalium', and that this is a matter of human choice (1.31). The eternal law is inclusive of the temporal; since the eternal law is more demanding than the temporal law, those who obey the eternal law obey the temporal law *a fortiori* (1.31). The practical relation of the two laws is made clearer at 1.33, by Augustine's first allusion to his doctrine of use: the things of the earth are neither good nor bad in the way that humans are, but they can be put to both good and bad uses by humans: '...cum praesertim videas et igne bene uti medicum, et pane scelerate veneficum'.

(Does the agent *use his will*, as well as all these other things? Should we translate 'voluntate' as 'voluntarily' or as 'by the will'? See below on 2.3, 2.48-54, 3.2-3, 3.47-50.)

From this a redefinition of evil appears. It is evident that there are two classes of men obedient to two different kinds of law because they pursue two different kinds of object; and 'quid autem quisque sectandum et amplectandum eligat, in voluntate esse positum'. Evil, then, is this deliberate choice of the earthly at the expense of the eternal:

'Licet... considerare, utrum sit aliud male facere, quam neglectis rebus aeternis, quibus per seipsam mens fruitur et per seipsam percipit, et quas amans amittere non potest, temporalia et quae per corpus hominis partem vilissimam sentiuntur et nunquam esse certa possunt, quasi magna et miranda sectari. Nam hoc uno genere omnia malefacta, id est peccata, mihi videntur includi.' (1.34)

1.35: *Conclusion.* Evodius agrees, and concludes that both the question of the nature of evil-doing and that of its cause have been answered: the nature of evil-doing is epitomised in the rejection of the eternal for the temporal, and the cause of evil-doing is human free will. Book One has argued that evil doing is cupiditas of a kind not necessarily punished by human law; and that the lawful life is the life of order, which is the life of the rule of reason in the individual; and that the life of reason is the life of the good will; and that the life of the good will is the life of the cardinal virtues and of happiness. By this series of identifications we are to be convinced of the possibility of the good life for humans, and of the doctrine that their rejection of the good life can only be by their own free choice or 'liberum arbitrium'.

3. Book Two

Summary of Contents:

- 2.1-4: If free will causes sin, why did God give us it?
- 2.5-6: Knowledge and belief.
- 2.7: Three questions proposed:
- 2.7-41: (a) How do we know that God exists?
- 2.42-47: (b) Do all good things come from him?
- 2.47-52: (c) Is free will a good thing?
- 2.7-14: Sensus corporis, sensus interior, ratio, and God.
- 2.15-28: The public and the private worlds 'seen' by these three faculties.
- 2.29-38: Wisdom, number and objective truth.
- 2.39-45: Number as the λόγος and formal cause of the world, and of all created things.
- 2.46-47: Hence, all good things come from God,
- 2.48-49: including the will.
- 2.50-54: The doctrine of use.

2.1-4. The question of the second book is 'Utrum ipsum liberum arbitrium, quo peccandi facultatem habere convincimur, oportuerit nobis dari ab eo qui nos fecit?' (1.35). Augustine answers that the will is necessary for *any* kind of moral life at all, good or bad: 'Si enim homo... non posset, nisi vellet, recte facere, debuit habere liberam voluntatem, sine qua recte facere non posset' (2.3). There could be no justice in God's punishing the natural use of the will; so if some uses of the will are

punished, it follows that they must be unnatural:

'...si quis *ea* [*sc. voluntate*] *usus fuerit* ad peccandum, divinitus in eum vindicatur. Quod iniuste fieret, si non solum ut recte viveretur, sed etiam ut peccaretur, libera esset voluntas. Quomodo enim iuste vindicaretur in eum, qui ad hanc rem *usus esset voluntate*, ad quam rem data est?' (2.3)

My italics mark the innovation that at this passage the will is first spoken of by Augustine as a thing *used* for good or for evil, as well as being the means of choice between good and bad use of other goods. I will first consider this.

If the will is used, then what is it used *by*? If Augustine's reply to this is 'By the agent', then the question comes: 'Why this apparent distinction between choices of good or bad things made by the will, and choices of good or bad uses of the will made by the agent? What sense can be given to the second idea which is not already given in the first?'. Accordingly, some Augustinian scholars, notably John Rist, have argued that

'*Voluntas* is not a decision-making faculty of the individual, as subsequent philosophy might lead us to suppose, but the individual himself. Hence it can be good or bad. Furthermore it is the basic core of the human person.'

(Rist (1969))

Rist claims that Augustine's view is that to have a good will just is to be a good person (Rist cites dGLA 2.4 and cIOP 1.101). Rist is evidently right, but his comment brings out the oddness of Augustine's talk here- and elsewhere- of the will as something which can be used (how do I *use* 'the basic core of my person'?). Rist suggests a possible smoothing of the roughnesses and oddities when he says that 'The word *voluntas*... denotes the human *psyche* in its role as a moral agent' (op.cit, p.421). There are many such *roles* which we find ourselves in, or *aspects* of human life which we experience, which are too particular to be identified with the *whole* of human experience, yet too pervasive to be plausibly seen as dependent on single organs/limbs, even ghostly organs/limbs: emotion, language, reason, aesthetic or moral responsiveness.

In context, at any rate, it is plain enough why Augustine suddenly begins to talk about the *use* of the will. It is because he is considering the will *as a gift from God* in dLA Bk.2. It is clear that gifts are things which the recipient may use; and one of God's gifts to us humans, in creating us the way he did, was free will.

But, it might be objected, the whole idea rests on a false analogy. For free will is not just another of God's gifts to humans, like legs, or perception, or imagination, to be used well or ill. Supposing that it is is like supposing that existence is a predicate just like other predicates. We should not say that God gave some created object yellowness, and curvedness, and

edibility, and *also* (as it happens) existence too. Nor (if 'free will' means 'power of choice to use things well or ill') should we say that God gave us, to use well or ill by our own power of choice, legs, perception, imagination, *and* free will.

Augustine's doctrine of the reflexivity of the will, on any reading, must involve the affirmation that, actually, you *can* properly say that. It is perfectly true (he may concede) that free will is not just another gift which is used like all the others. But (as he explicitly argues: 1.25, 2.3) this does not mean that it is not a gift-which-is-used at all. On the contrary: it is only because *this* gift is in use, that any of the others at all can be. The doctrine of the reflexivity of the will which Augustine adopts is just the claim that the will *is* used. By what? By itself: choices of how to use the will are themselves choices made by the will.

Does this condemn Augustine to Ryle's vicious regress? Not necessarily. On one reading of some of his remarks which I have been considering, Augustine could be committed to the claim that, behind *every* choice made by the will, there stands another choice made by the will about how to use the will. And indeed, in his extremer statements of the reflexivity theme, Augustine may commit himself to this kind of idea.

But there is also a more moderate understanding of reflexivity to hand. Augustine could make a distinction between first-order choice (of simple intensional objects) and

second-order choice (where the intensional object is a characterisation of what kind of first-order choices I would like to be disposed to make). Plainly, only some choices are second-order. Equally plainly, if I am capable of controlling what I choose at this level, I have the power to choose (second order) what to choose (first order). This does not necessarily mean that there must be a third, and fourth, and fifth,... order; only that 'voluntas' as Augustine conceives it is patient of these kinds of distinctions of order, which it may sometimes, in a non-regressive manner, be useful to employ in analysis.

Augustine, then, points out that freedom of choice can be misused. Evodius' response is one borrowed by John Mackie: 'Si ad recte faciendum [voluntas] data est, non debuerit ad peccandum posse converti' [my italics]. Evodius compares justice, also given to humans so that they may live well: *this* cannot be used for ill purposes, so how is it that free will can be?

Augustine responds by proposing the argument of 2.7-54, centred on these three questions:

- (i) quomodo manifestum est Deum esse;
- (ii) utrum ab illo sint quaecumque in quantumque sint bona;
- (iii) utrum in bonis numeranda sit voluntas libera.

(i-ii) Augustine's argument to establish the first two points is as follows. It is clear that God exists because of the objectivity of the order in the world which humans are capable of

grasping. This order and truth, as the object of human reason, is superior to that reason (2.37). Therefore, since there is nothing superior to reason which is not God (2.14), this order and truth which reason perceives is to be identified with the presence of God (2.39). Now this order and truth which has been taken to show God's presence is both good and all-pervasive in creation (2.31). Therefore, we may indeed say that whatever is good, insofar as it *is* good, is from God (2.46). Hence, if (and insofar as) free will is a good, it *must* be from God.

(iii) Augustine then (2.47-54) has to show that free will a good thing, yet not the kind of good that cannot be turned to evil. First, some comments on 2.7-47.

2.7-14: Sensus corporis, sensus interior, ratio, and God. At 2.7-14 Augustine gives us, in small scope, a philosophy of mind, the importance of which for his philosophy of action (and hence for his theory of the will) will become obvious as soon as we go into it. The key to this theory is again the human mind's essentially reflexive nature, which Augustine establishes at the very first step. The point has already been made at 1.16 that awareness of one's *own* existence is a grounding certainty. It is repeated here more clearly:

'Utrum tu ipse sis? An tu fortasse metuis, ne in hac interrogatione fallaris, cum utique si non esses, falli omnino non posses?' (2.7)

That I am a self-conscious living existent can, it is claimed, be established by simple introspection. And from *that*, it follows *a fortiori* that I am both a living existent, and an existent.

This is the theme of reflexiveness again, and the *motif* is repeated in Augustine's analysis of perception. The bodily senses are shared by humans and animals, and 'sentire' is accordingly seen by Augustine as on the same level in his 'scala naturae' as 'vivere' (1.16, 2.8). But a question arises about how the interrelation of the bodily senses is achieved. For none of the senses themselves can be the means whereby we distinguish the reports of the different senses from each other. This, Augustine argues, requires us to suppose that there is a 'sensus intus' which discriminates between them. (Is this 'sensus intus' a descendent of Aristotle's κοῖνη αἴσθησις?)

Augustine's question is: How do we *perceive* the difference between a visual sense-datum, and (say) an olfactory one? His answer is: By means of the 'sensus intus'. For this is the sense which has, as its objects of perception, not just first order sense data *but the first order senses too*: 'Sensum illum interiorem non ea tantum sentire, quae acceperit a quinque sensibus corporis, sed etiam ipsos [sensus] ab eo sentiri' (1.10).

But the 'sensus interior' is itself subordinate to something else: namely reason. And here comes the reflexivity again: Reason

is distinguished from the 'sensus interior' in that reason not only perceives both the first order sense data, and the first order senses, as the 'sensus interior' does; and not only perceives also the second order sense, the 'sensus interior', as we might expect; reason also perceives itself, reflexively, as well as everything else, non-reflexively.

'Manifesta enim sunt, sensu corporis sentiri corporalia; eundem autem sensum hoc eodem sensu non posse sentiri; sensu autem interiore et corporalia per sensum corporis sentiri, et ipsum corporis sensum; ratione vero et illa omnia, et eandem ipsam notam fieri, et scientia contineri.' (2.10)

This suggests an emendation of a standard reading in a crabbed passage in 2.9. This is Augustine's account of how sense data are mediated to the human reason, as it stands in Thonnard's text (which is Migne's, vol. 33, 1221-1303):

'Quid? Ista ratio posset haec quattuor discernere ab invicem et definitionibus terminare, nisi ad eam referretur et color per oculorum sensum, et ipse [oculorum sensus] rursus per illum interiorem [sensem] qui ei praesidet, et idem interior [sensus] per seipsum, si tamen iam nihil aliud interpositum est?'

Here, I think, consistency with the rest of Augustine's exposition requires the reading *seipsam* [sc. rationem] for Migne's *seipsum* [sc. interiorem sensum] in the last line. For, from the rest of Augustine's argument, it is plain that the

reflexivity of which I am making so much does not occur at the level of the 'sensus interior', but at the next level up, that of the reason. The 'sensus interior' cannot perceive *itself* any more than any other sense can; therefore it does not function 'per seipsum', but is one of the intermediaries in the chain: sense data > 'sensus exteriores' > 'sensus interior' > reason. But reason *can* perceive, or rather have awareness of, itself ('eandem ipsam notam fieri'), and it is no kind of intermediary; therefore, reason does function 'per seipsum'. Given this emendation, we may translate:

existence

1101. 'How could the reason distinguish these four [sense data, senses, the interior sense, and reason] from each other, and define their limits, unless what was conveyed to it was all of these?— (i) the colour, by the sense of the eyes; (ii) that same sense [of sight], by the interior sense which is set over it; and (iii) that interior sense, *by reason itself?*'

can only

1102. Of course, all of this is not without its problems. The very idea of a 'sensus interior' may seem highly suspect to modern minds: what does this wheel turn? Augustine might also be accused, with some justice, of using 'ratio' as a *deus ex machina* to solve, or rather shelve, problems which are caused simply by the inadequacies of his account of perception. If the upshot of his story is simply that reason can 'make known to itself' both itself and all the yieldings of every subordinate form of awareness, then what is the point of all the preliminaries to this not very explanatory remark? And isn't Augustine just

committing a category blunder when he talks about *perceiving* senses?

I will not defend him from these charges here. My main aim, in this account, is simply to point out the very strong parallels between Augustine's use of reflexivity in his theory of action and in his theory of knowledge.

2.14-38: *Wisdom, number and objective truth.* This account of the mind is followed by the transition to the argument for God's existence (2.14-39). If there is something superior to the human mind, something eternal and unchangeable, this will be God. Now Objective Truth, grasped by the reason, is superior to that reason (35). Therefore God is Objective Truth.

Once again the *scala naturae* plays a part in this argument: Augustine proceeds from the objects of outer sense, in which we can only have a certain degree of community since they are expendable, to the objects of reason, which are not expendable and hence capable of being absolutely shared. Your part in seeing some physical object is never quite the same as mine. If you were ever to look at Botticelli's *Primavera* from exactly the same viewpoint as that from which I looked at it, it would follow from that that I was *not*, at that moment, looking at *Primavera* from that viewpoint. My looking at *Primavera* at a given time, from a given place, is not compatible with anybody else's looking at it at the same time and place— as anyone who visits the Uffizi in August will quickly discover. But your part in grasping the kind

of Objective Truth that Augustine has in mind is completely independent of mine, and has no tendency to diminish my ability to grasp it. Augustine concludes that Objective Truth is the most unchangeable and eternal object of contemplation because, in this sense, it is the most absolutely *public* object of contemplation (2.19). Further, the perfect form of Objectivity in this sense seems to be identical with mathematical truth (2.32: 'cum eadem sit'), which is the object of pure reason:

'Qua phantasia vel phantasmate tam certa veritas numeri per innumerabilia tam fidenter [novimus], nisi in luce interiore conspicitur, quam corporalis sensus ignorat?' (2.23)

From this Augustine draws a conclusion of importance to his (thus far unexpressed) conception of practical reason. Since *objective truth*, as defined, is unitary, it follows (2.25-27) that *wisdom* is unitary too: 'Singulas quasque suas arbitraris singulos quosque homines habere sapientias?' (2.25). The Latin 'sapientia' is usually more suggestive of the Greek σοφία than φρόνησις; but that Augustine means principally to discuss *practical wisdom* is plain at once from Evodius' response to his question, which concerns the goals of action.

In reply Evodius admits the unity of wisdom, but adds, 'Video quippe varie videri hominibus, quid fiat dicaturve sapienter' (2.25). He cites the military life, the contemplative life, and the political life, as providing apparently incommensurable τέλη at which reason might aim, and in pursuit of which wisdom might

be thought to be best exercised. Augustine's reply to this lays bare the heart of his thinking about practical wisdom. And once again the crucial links between Augustine's cognitive and conative philosophies become apparent.

Augustine starts by defining 'sapientia': 'Num aliam putas esse sapientiam nisi veritatem, in qua cernitur et tenetur summum bonum?' (2.26). Practical wisdom is the possession of the highest good in virtue of understanding a certain kind of truth, *practical truth*. It follows that without practical truth, there is no practical wisdom. What kind of truth is this? It is the truth about what is *really* the 'summum bonum', independently of and antecedently to the various views which may be held:

'Nam illi omnes quos commemorasti diversa sectantes, bonum appetunt et malum fugiunt; sed propterea diversa sectantur, quod aliud alii videtur bonum. Quisquis ergo appetit quod appetendum non erat, tametsi id non appeteret nisi ei videretur bonum, errat tamen... In quantum igitur omnes homines appetunt vitam beatam, non errant. In quantum autem quisque non eam tenet vitae viam quae ducit ad beatitudinem, cum se fateatur et profiteatur nolle nisi ad beatitudinem pervenire, intantum errat... Et quanto magis in via vitae quis errat, tanto minus sapit.' (2.26)

Augustine's practical truth turns out to be exactly the same in content as his theoretical truth. It is natural to all men to seek the (apparent) good and flee the (apparent) evil (cp. 1.30, 'nam hoc volunt etiam mali'). But what more is necessary to the

good life is exactly the same knowledge of transcendent and objective reality as forms the end of enquiry. I said that, for Augustine, wisdom is as unitary as the objective truth which it contemplates. It follows from this that, in fact, there is no neat division between practical and theoretical wisdom in his thought. Just as, for Augustine, talk of volition and reason relates to the same human person in the different roles of agent and percipient, so also, for him, *practical* reason is only theoretical reason put to work, and *theoretical* reason is only practical reason in contemplative mode. The first principles of practical and theoretical reason are, in his view, one and the same.

Evodius is quite capable of drawing the right conclusion from this: 'Si summum bonum omnibus unum est, oportet etiam veritatem in qua cernitur et tenetur, id est sapientiam, omnibus unam esse communem' (2.27). Since practical wisdom is identical with theoretical wisdom, and theoretical wisdom is the unitary possession of objective truth, it follows that there is only one genuine practical wisdom. The form of statement on the grounds of which an ascription of the predicate 'practically wise' would be justified is: 'X has a good reason to ϕ '. And such statements have, for Augustine, quite determinate and objective truth-values, irrespective of X's actual reasons for ϕ ing.

Evodius admits all this as plausible, but is still impressed by the great variety of objects actually treated as 'summa bona': 'Dubito sane, quia diversos diversis rebus gaudere video tanquam

summum bonum suis' (2.27). Augustine argues that the fact that there is a great variety of objects considered as goods (or even as *the* good) within some limited area of choice does not entail any disturbing conclusions about the incompatibility or incommensurability of those goods. The one concept of goodness is as common to all these choices as is the sunlight to the beauties that we see:

Text: 'Etiam si multa sunt bona eaque diversa, e quibus eligat quisque quod volet, idque videndo et tenendo ad fruendum summum sibi bonum recte vereque constituat; fieri tamen potest ut lux ipsa sapientiae, in qua haec videri et teneri possunt, omnibus sapientibus sit una communis.' (2.27)

Burleigh (1953)'s translation of part of this runs: 'there are many diverse good things from among which each may choose what he likes, and... may rightly and truly constitute it his own chief good' (Burleigh 1953, p.152). This translation, which suggests that what Augustine means by this is that all choices of a 'summum bonum' for a whole life imply the same criterion of choice, ought to puzzle us. For Augustine is in the middle of arguing that there is only *one* right and true choice of 'summum bonum'. What Augustine's references to different aesthetic tastes in scenery rather suggest is that he is arguing that different people can hold quite different 'bona' as the 'summa' for *smaller-scale* areas of judgement than the judgement of what is 'summum' for the whole life. If so (the evidence, like the argument, is unclear), then Augustine is introducing something

like a distinction between *practical* and *technical* reasoning: within these smaller (technical) contexts of choice of the good, there can perfectly well be different views about what is the best, without this fact undermining the unitary nature of the ethical choice to be made regarding the *whole* of a human life. Augustine thinks there is one best *life* for humans, but bids us consider, for example, how there is not and need not be one greatest *painting*; rather, there is a fuzzy-edged set, The Great Paintings. Augustine is simply pointing out to Evodius that this kind of technical indeterminacy may be what is making him suspect the existence of real practical indeterminacy- which of course Augustine says does not exist. The criteria of technical rationality will be subjective in a way that those of practical rationality are not.

From considering the practical side of his unitary 'sapientia', Augustine moves to a more contemplative mode (2.30-36). What is relevant for this discussion is the reaffirmation that it is exactly this contemplated truth which is the first principle of action, and the point at which a context of commensurability is found between all smaller technical contexts:

'Haec enim veritas ostendit omnia bona, quae vera sunt, quae sibi pro suo captu intelligentes homines, vel singula, vel plura eligunt, quibus fruuntur... Sic fortis acies mentis et vegeta cum multa vera et incommutabilia certa ratione conspexerit, dirigit se in ipsam veritatem, qua cuncta monstrantur.' (2.36)

Augustine brings together two earlier claims to support his argument about the identity of practical and theoretical truth. At 1.10 ff. he argued that the essence of wrongdoing was the pursuit of what one can lose against one's will; at 2.19 ff., that the contemplation of physical goods can be frustrated by their expendability, by the fact that one contemplator of such goods is, at least to some degree, necessarily in competition with other contemplators. Putting these two points together (2.37) he gets the claim that the pursuit of truth is the only virtuous pursuit, for only truth is absolutely common and equally present to all its pursuers- provided, of course, that we simply will to pursue it.

'At illa veritatis et sapientiae puchritudo, tantum adsit perseverans voluntas fruendi, nec multitudine audientium constipata secludit venientes, nec peragitur tempore, nec migrat locis, nec nocte intercipitur, nec umbra intercluditur, nec sensibus corporis subiacet. De toto mundo ad se conversis qui diligunt eam, omnibus proxima est, omnibus sempiterna; nullo loco est, nusquam deest; foris admonet, intus docet; cernentes se commutat omnes in melius, a nullo in deterius commutatur; nullus de illa iudicat, nullus sine illa iudicat bene.' (2.38)

And this objective and absolute truth, both as seen in the world (2.42) and as seen in the canons of practical choice which we employ (2.41), turns out, again, to be identifiable with the presence of God (2.43, 2.39). Also with Number (2.42), the presence of which throughout creation is also evidence of the

good giving of God (2.45). So at length, with answers of a kind proposed to the first two questions of 2.7 ('How do we see that God exists?'; 'Do all good things come from God?'), Evodius comes back to the third question: 'Is free will a good gift?' (2.47).

2.46-47: *All good things from God.* Augustine's first response is to claim that he has already answered this. For it has already been agreed (2.3) that humans cannot act well without free will; which itself makes free will a good gift. The problem with this, of course, is that it also establishes that free will is a *bad* gift, *pari passu*. Augustine needs a further argument.

For this, he distinguishes three kinds of goods: 'magna bona', 'media bona', and 'minima bona' (2.50). This classification is arrived at by asking two questions about any putative good G:

- (1) Is G's possession necessary to the good life?
- (2) Is it impossible to misuse G?

Three kinds of good? -Or four? These questions, taken together, generate four answers (YY, YN, NY, NN), not three. Justice, like all the virtues, is an example of the highest kind of good (YY): it is essential to the good life, and it cannot be misused (2.50). The parts of the human body are examples of the lowest kind of good (NN):

'AUG. Iam ergo tu negabis luscum hominem recte posse vivere?

EV. Absit tam immanis amentia!

AUG. ...Quo amisso tamen ad recte vivendum non impeditur.' (2.49)

And free will is a 'medium bonum'. Without it none can live aright (2.49), yet it can be misused (2.48). So it is a 'medium bonum' of the (YN) type.

(In passing: The (NY) type of 'medium bonum' is here ignored by Augustine. But if my interpretation of 2.26-27 is right, an interesting place in Augustine's philosophy can be found for this type. 'Media bona' which cannot be misused, yet which are not essential to the good life, will (I suggest) characteristically be the kind of aesthetic and other experiences which Augustine draws our attention to at 2.27. As I have noted, these are the kind of things which provide the smaller-scale 'summa bona' which are the τέλη of particular technical excellences. Evidently I need not have, e.g., a profound vision of the dramatically sublime to be a good man, nor indeed engage in the dramatic τέχνη at all (of which, officially, Augustine of course disapproved). But if I do so engage, then the θεωρία of that technical 'sumum bonum' is itself a good. I can misuse my own technical excellences; but I cannot misuse the *standards* of the technical excellences.)

The difference between a 'magnum bonum', like justice, and a 'medium bonum', like free will, is in a sense just a grammatical one. It is built into the meaning of 'justice'- of any virtue- that one cannot act so as to misuse it. To 'use justice' at all,

if we care for this locution, is to use it for a good end. 'To use (the virtue of) justice for a bad end' could only mean 'not to use justice, but to use *injustice*'. This point, of course, applies equally to 'good will' (as opposed to 'free will'). To use good will for a bad end is not to use good will at all, but bad will. Good will, in short, is a 'magnum bonum' like the virtues, not a 'medium bonum' like free will. (Thus Augustine's 'bona voluntas' cannot possibly be identified in sense with his 'libera voluntas'.)

A little reflection on Augustine's threefold (or rather, I have argued, fourfold) division of the goods shows a striking point. There are indefinitely many goods of the (YY) type— as many as the virtues which one might want to enumerate. There are also indefinitely many (NN) goods: as many as there are physical objects, perhaps. And if I am right, Augustine himself can be seen as arguing in effect that there are indefinitely many (NY) goods too— as many as there are τεχνάϊ. But does not Augustine's own argument tend to the conclusion that *there is only one (YN) good, free will itself*? In general, whatever is necessary for the good life, like the virtues and good will, is so internal to the definition of the good life that it is logically impossible for it to be misused. Thus, to have a good will *is* to be living the good life (1.10). And whatever can be misused is not generally such as to be internal to the good life as Augustine defines it. Thus, nothing which we can lose against our will is worthy to be considered a good (or better, *the good*) (1.10). The only exception to these two rules is free will itself. This alone is a

necessary part of the good life; yet can be misused.

2.50-54: *The doctrine of use.* This must be partly because to speak of the will at all is to speak of *use*. It is to speak of a capacity which stands alongside and outside all other capacities, and yet remains a capacity itself; indeed the capacity *par excellence*, for without it there are, in effect, no other capacities. This is what I have been urging by my insistence on the notion of reflexivity; and this is the point which Evodius is brought to understand at 2.51:

'EV. Quomodo et ipsa [voluntas libera] inter illa quibus utimur numeranda sit?

AUG. Quomodo omnia quae ad scientiam cognoscimus, ratione cognoscimus, et tamen etiam ipsa ratio inter illa numeratur quae ratione cognoscimus... ut quodammodo se ipsa utatur voluntas quae utitur ceteris, sicut seipsam cognoscit ratio, quae cognoscit et cetera.'

This is a clear statement of the doctrine of the reflexive use of the will by the will. It also points up nicely my claim that there is a strong parallelism between Augustine's conceptions of the will and of the reason.

The happy life, then, is to share in the 'commune bonum', not the 'proprium bonum'; and whoever does so is happy, but whoever does not, sins (2.52-3). Not that even the objects of desire of the worst humans are bad in themselves; but they are *private*

goods, *changeable* goods (2.53). Augustine now states baldly (2.53) that this 'aversio ab incommutabili bono, et conversio ad mutabilia bona' is, 'quoniam non cogimur', voluntary. But this needs to be argued (2.54): what is the cause of this 'aversio'? Augustine's standard reply to this is 'nescio'. By this he does not mean that he is ignorant, but there is nothing there to be known. This is a version of Augustine's rather Protean argument that evil is not a substance but a privation of good: 'Detracto penitus omni bono, omnino nihil remanebit' (2.54). Here the application is not so much to substance as to the movement of the will: 'Motus ergo ille aversionis... defectivus motus est, omnis autem defectus ex nihilo est' (2.54). This might mean that the 'aversio' is not compelled by main force; but doesn't it show this at the price of making it out to be a completely undetermined (i.e., random) 'motus'? Augustine thinks that, even if it is random, at any rate we do not have to let the 'aversio' happen in us:

'Qui tamen defectus quoniam est voluntarius, in nostra est positus potestate. Si enim times illum, oportet ut nolis; si autem nolis, non erit. Quid ergo securius quam esse in ea vita, ubi non possit tibi evenire quod non vis?' (2.54)

I can *choose* to adopt this kind of 'aversio', it does not necessarily come on me like a mad fit; I can also choose to avoid it (cp. dVR 27). But, of course, once fallen, it is not so easy for mankind to rise again: what is needed is grace, the outstretched right hand of God- an important remark which makes

it clear that, even at this early stage, there is another side to Augustine's apparently rather simple-minded confidence that the good life can be obtained simply by willing.

How far is this analysis of the causes of sin consistent with, and understandable in the light of, what has been said already about the reflexivity of the will? Some hints at an answer to that have been given already; a fuller statement can best be made in the context of Book Three, where the issue of the origin of sin is discussed more adequately (3.37-77). To which I now turn.

4. Book Three

Summary of contents:

- 3.1: How does the agent fall into wrongdoing?
- 3.2-3: Natural necessity contrasted with human liberty; the doctrine of use again. 'Neque culpandus homo esset cum ad inferiora *detorquet quasi quemdam cardinem voluntatis*.'
- 3.4-11: Foreknowledge and predestination.
- 3.12-18: The perfection of the universe in its variety and its balance.
- 3.19-25: Compulsion, misery, suicide and the goodness of existence *qua* existence.
- 3.26-37: Cosmic balance and sin.
- 3.38, 42: Praise and blame: 'Nullius autem vituperatur vitium, nisi cuius natura laudatur'.

- 3.39-41: How corruption happens.
- 3.43-46: Vindication of God's justice.
- 3.47-50: The causes of the will.
- 3.51-55, 64-65: Ignorance and voluntariness.
- 3.56-63: The origin of souls and original sin.
- 3.66-69: Infant and animal suffering.
- 3.70-75: Wisdom, folly and perception.
- 3.76-77: God and the created mind.

3.1: *How does the agent fall into wrongdoing?* 'What we should call psychological compulsions are not compulsions for Augustine. They are simply the individual working out his own nature' (Rist 1969, p.422). If this is right, then it seems odd that, at the beginning of dLA Bk.3, Augustine makes it quite plain that in his view 'Si enim natura vel necessitate iste motus [peccandi] existit, culpabilis esse nullo pacto potest' (3.1). That this is his view appears in a good number of different places in his early writings, and not just the dLA.

As Evodius is made to say: 'Nec quod naturaliter movetur recte vituperari potest; quia etiamsi ad perniciem movetur, naturae tamen suae necessitate compellitur' (3.2). In which case, as Augustine points out, the ascriptions of praise and blame which Evodius is so strongly inclined to make of the corrupt will (3.1, 3.2) would be senseless. If free will is such that it is *bound* to go wrong, of necessity, then its movement is not free; or at any rate not responsible. In fact the stone (as conceived

in the Aristotelian physics of Augustine's day) provides a remarkably close analogy to what such a will would be like. A stone in mid air has no external force impressed on it. Nothing outside it forces it to go any one way rather than any other. The cause of its fall is within it: simply (in Rist's phrase) 'the working out of its own nature', the stone's own earthward propensity.

But in fact, Augustine qualifies, while it is true that the soul's 'motus aversionis' is, like the stone's, a '*proprius motus*', there is this difference: the soul can choose *not* to start falling:

'In potestate non habet lapis cohibere motum quo fertur inferius; animus vero dum vult, non ita movetur, ut superioribus desertis inferiora diligat. Et ideo lapidi naturalis est ille motus, animo vero iste voluntarius.' (3.2)

It is important to notice what Augustine does not say here: namely, he does not say that the soul can arrest its fall once begun. Nor does he say that a soul which so fell, and was unable to rise again, would not be culpable for this inability and for anything which resulted from it. Such differences in doctrine from what Augustine was later to oppose in the Pelagians are to be borne in mind. On the other hand, almost in the same breath he makes it sound as if that fall had not already occurred, was something within each individual's present control:

118. 'Omnisque de hac re disciplina utilis ad id valeat, ut eo motu improbato atque cohibito, voluntatem nostram ad fruendum sempiterno bono, a lapsu temporalium convertamus.' (3.2)

119. Evodius (3.3) sums up the findings of the dialogue so far. The will is an immediately present item of self-awareness, and the possession most truly my own; it pursues what brings joy; it is necessary for the good life, and its use makes us culpable for bad living and answerable for all our actions; its existence and freedom is implied by divine admonition and command. Without it there would be no responsibility; which brings us to the extraordinary phrase of 3.3:

120. 'Neque culpandus homo esset cum ad inferiora detorquet *quasi quendam cardinem voluntatis*.'

121. What is Augustine's meaning here? Burleigh (1953, p.172) has '...Nor blameworthy, when he turns to lower objects, using his will like a hinge'; Thonnard (1941, p.329) reads '...Ni coupable, lorsque pivotant, pour ainsi dire, sur elle-même, elle se détourne vers les biens inférieurs'. But does 'cardinem voluntatis' mean, as Burleigh implies, that the will itself is the hinge; or is it that there is a part of the will which is used as a hinge? In either case, the familiar question arises: What is it that does the pivoting of the will or uses the will as a hinge? To this the only possible answer seems to be: The will itself. So perhaps the correct answer is that Augustine means that the will is both active and passive simultaneously in this

transaction: plays the roles both of mover and of hinge, is both the pivoter and the pivoted.

3.4-11: *Foreknowledge and predestination*. Evodius asks: 'Quoniam peccatum esse praesciverat, necesse erat id fieri, quod futurum esse praesciebat Deus. Quomodo est igitur voluntas libera ubi tam inevitabilis apparet necessitas?' (3.4). In response Augustine makes two gambits: first, he remarks that what God foreknows is that I act *freely* (3.8); which seems to suggest that Augustine's view is like Luis de Molina's, that, necessarily, I act freely. Second, he remarks that human prescience in itself has no tendency to limit freedom of action at all; so why should it be different with Divine prescience (3.10)? By 3.11 the distinction between compulsion and foreknowledge is clear enough for Augustine to ask: 'Cur ergo non vindicet iustus, quae fieri non cogit praescius?'. He considers the problem solved from this point on. (It was not solved, of course: consider Molina (1588) and the huge recent literature on the subject, from Pike (1964) and Plantinga (1976) onwards.)

3.12-18: *The perfection of the universe*. At 3.12 Augustine turns to the question of 'Quomodo non Creatori deputandum sit quidquid in eius creatura fieri necesse est?'. This question (of theodicy rather than of freedom) occupies much of Augustine's attention up to 3.46. The central points Augustine makes are (i) that the existence of what he (3.31) calls 'naturae'- substances- is a good in itself, and hence (?) so is the existence of as many different 'naturae' as possible (3.16, 22-24). And (ii) that it is

a matter of indifference for God's glory whether humans sin or not (3.31). For if they do not, God's justice is shown in rewarding them; but if they do, His justice is shown in punishing them.

'Miseriam quam doles, ad id quoque valere cognoscas, ut universitatis perfectioni nec illae desint animae, quae miserae fieri debuerunt, quia peccatrices esse voluerunt.' (3.25)

Augustine remarks (3.15) that 'Est excellentior creatura quae libera voluntate peccat, quam quae propterea non peccat, quia non habet liberam voluntatem'. Augustine's theodicy involves a double notion of goodness. Natures can be better or worse according to their place in the *scala naturae*. But they can also be better or worse according to how successfully they play that role in the *scala*, how truly they are what they truly are or ought to be. Failing in this is a 'privatio boni', by which Augustine seems to mean indifferently a privation of *good*, or of *substance*, or of *existence*.

3.19-25: *Compulsion and voluntary misery*. Augustine offers, *en passant*, some remarks on the involuntary (3.19), designed, in context, to show that no one is miserable unjustly. Either one is in another's power, or in one's own; either one is miserable or not. If one is in one's own power, and not miserable, then there is no problem; if miserable, then this is caused by one's own unjust self-government. Or if one is in another's power, this other must either be someone who is able to overcome the soul, or

someone who is not so able. But only God is able to overcome the soul (2.14), and whatever one experiences from *Him* is of course justly dealt. And if one is in the power of someone who is not able to overcome the soul, then this can only be because one chose so to be; for which choice one is answerable. For (3.29) there are 'duae origines peccatorum, una spontanea cognitione, alia persuasione alterius'; but 'utrumque voluntarium est'. This argument is reiterated at 3.38-40.

3.21 gives us a different version of the kind of reflexive principle of virtue expounded in the first book. Compare (1.26) 'Quid enim tam in voluntate, quam ipsa voluntas sita est?' with (3.21) 'Si vis itaque miseriam fugere, ama in te hoc ipsum, quia esse vis... quanto amplius esse amaveris, tanto amplius vitam aeternam desiderabis'. To love existing is to love life, according to Augustine.

At 3.32-33 Augustine seems at first sight to be teaching something like double predestination. He divides free created natures into the higher and the lower, the 'in virtute ac iustitia permansuras' and the 'peccaturas' (3.32). By this he does not seem to mean a division between, say, angels and humans, for evidently some angels fell. Rather he seems to be dividing the elect (human and angelic) from the reprobate (ditto). He tells us (3.33) that the universe's order would be impaired if the higher sort either did not exist, or if they sinned; but it would only be impaired by the lower sort if they did not exist. And most significantly of all,

'[Sublimiori naturae] data est potentia omnia continendi officio proprio, quod rerum ordini deesse non possit; nec ideo in bona voluntate permanet, quia hoc accepit officium; sed ideo accepit, quoniam ab illo qui dedit permansura praevisum est.'

(3.33)

Here there is no mention of any direct predestination, only of prescience. But this prescience is made the ground for God's giving of an 'officium omnia continendi', not given to other natures, which is given on the sole ground that God's prescience distinguishes the higher kind of nature from the lower. The idea seems to be that God's grace comes only to those whom He foresees will accept it.

3.34: *Body and Soul*. Augustine explicitly says that the sinful soul is not ruled only by its own volitions acting on a strong body, like the sinless soul, but also by the laws of physical determinism acting on a weak body:

'In corporibus autem inferioribus atque mortalibus post peccatum, anima ordinata regit corpus suum, non omnimodo pro arbitrio, sed sicut leges universitatis sinunt.'

(3.34)

However, he does not seem to build anything on this—surprisingly, perhaps, from a modern viewpoint.

3.47–50: *The causes of the will*. Augustine here lays out his reasons for affirming a doctrine which will be of great

importance in the *City of God*, and which is central to my interpretation of his thought: that there is no cause of an evil will.

Note that I say 'of an *evil* will'. For Augustine is not saying that *every* 'voluntas' is uncaused, as (e.g.) Thonnard evidently takes him to be (Thonnard's chapter heading for dLA 3.48 is 'Il n'y a pas de cause au delà de la volonté libre': Thonnard (1941), p.416). But note carefully the transition from Evodius' question to Augustine's answer:

[Evodius:] 'Ego enim causam quaero ipsius voluntatis. Non enim sine causa nunquam vult illa [creatura] peccare, nunquam ista non vult, quaedam vero aliquando vult, aliquando non vult, cum eiusdem generis omnes sint. Hoc solum enim mihi videre videor, non sine causa esse tripartitam voluntatem rationalis creaturae; sed quae causa sit, nescio.' (3.47)

Augustine's reply begins thus:

'Quoniam voluntas est causa *peccati*, tu autem causa ipsius voluntatis inquiris...' (3.48)

Evodius is asking about the causes of all three of his kinds of wills, the wholly good, the wholly bad, and the mixed. But Augustine's reply restricts itself to the willing of bad deeds. This, surely, implies that Augustine assents to at least part of what Evodius says: namely, that there must be a cause of the *good*

will's being as it is. In Ch.7 I will argue that the notion of practical wisdom- as, in fact, developed in the dLA itself- gives us a crucial part of Augustine's theory of what it is for a good will to have causes. Thus, if Augustine were to say at dLA 3.47-49 that the will in general has no cause, he would be contradicting not only his own opinions, but his own opinions in the same work.

The evil will does not have any cause, in the sense that the good will has a cause: this, *not* the claim that the 'voluntas' in general is subject to indeterminacy, is Augustine's point in dLA 3.48-49. We should not be misled by Augustine's use of 'causa voluntatis' into taking him to be making out a general claim. For the only 'voluntas' he is interested in here is that 'voluntas' which is the 'causa peccati'. Thus:

'Quae tandem esse poterit ante voluntatem causa voluntatis? Aut enim et ipsa voluntas est; et a radice ista voluntatis non receditur; aut non est voluntas; et peccatum nullum habet.'

(3.49)

What does it mean to say that (not 'voluntas' in general, but) 'improba voluntas' has no cause? More about that in Chs.8-9. For the moment I will simply point out the connection of this idea with the idea of 'privatio boni'. (Cp. 3.38 and 3.41, where Augustine tells us that whatever is a fault is contrary to the nature in which it is the fault; and that the fault is to be blamed precisely because the nature is to be praised.)

3.51-6: Augustine considers the relation of ignorance and incapacity to voluntariness. Fallen man acts wickedly either because 'non habet in potestate ut bonus sit', a state which is now second nature to him (the ice is thin here: if wrongdoing is almost natural to humans (3.52, 'violentia... quodammodo naturaliter inolevit') then they are barely responsible for it, 3.38). Or else he acts wickedly, by 'non videndo qualis esse debeat' (3.51). His state is attended everywhere by the conditions 'ignorantia' and 'difficultas' (3.52). (Is this ignorance, in Aristotle's terms, ignorance of moral principle or of facts?)

Not everything so done in ignorance or because one cannot help it is to be pardoned (3.51: 'Etiam per ignorantiam facta quaedam improbantur'). But on the other hand, Augustine describes these conditions as 'duo iste poenalia' (3.52), penal conditions imposed in punishment for the first sin of the first (and only, before Christ) free human:

'Illud quod ignorans quisque non recte fecit, et quod recte volens facere non potest, ideo dicuntur peccata, quia de peccato illo liberae voluntatis originem ducunt.' (3.54)

'Cum autem de libera voluntate recte faciendi loquimur, de illa scilicet in qua homo factus est loquimur.' (3.52)

The natural question now is: Is it just to punish us for Adam's sin? Augustine refuses to pursue this. He merely

reiterates his earlier arguments that God is necessarily just (3.51): 'Let them be silent, and stop murmuring against God' (3.53). In any case, he adds as an afterthought (3.53), it is fallen humanity's unresponsiveness to God which is culpable, not its mere fallenness.

Augustine holds, like Aristotle, that if an act is culpable, it is voluntary; and if it is voluntary, this means that the agent could have done otherwise:

'Ex conversione ad Deum, ut vinceret quisque supplicium quod origo eius ex aversione meruerat, non solum volentem non prohiberi, sed etiam adiuvari oportebat.' (3.55)

Even fallen humans, it is here quite clearly affirmed, are able to desire God. And when they do so, God's grace is on hand to aid their desire. Again, in 3.65 Augustine stresses that it is within human power to choose whether to become good or not.

In 3.66-70 he discusses, rather unsymathetically, the plight of those creatures which do not seem to have such powers of choice, infants and animals. Of infants Augustine suggests (3.68) that what is not yet sinful is not yet good either; so that its suffering is unimportant. Of animals he suggests (3.69) that their suffering may contribute to the harmony of creation.

Finally, in 3.71-77, Augustine discusses the objection raised by Evodius at 1.24: 'If man was created wise, then he could not

be so foolish as to succumb to vice; if he was created foolish, then this was God's fault, not his'. Augustine replies that man was created neither foolish nor wise:

'Tunc enim homo incipit aut stultus esse aut sapiens, cum iam posset, nisi negligeret, habere sapientiam, ut vitiosae stultitiae sit voluntas rea.' (3.71)

The character of a person who has not yet chosen either for good or for evil is as yet unformed; ascriptions of 'wise' and 'foolish' depend for their truth upon the existence of dispositions which have not yet been formed. 'Incipit homo praecepti esse capax, ex illo incipit posse peccare' (3.72).

But (3.74) since the agent necessarily acts on what information is available to her, and the information available to Eve included the Serpent's voice, how could she avoid seduction? But Augustine responds (3.75) that this is not all the information present to the agent. Besides that temptation, there is also the awareness (*intentio*) of God's presence, and the reflexive awareness of oneself: 'Subiacet ergo intentioni animi prius ipse animus, unde nos etiam vivere sentimus: deinde corpus quod administrat'. All of these parts of our natural awareness go against the temptation of the Devil, which- Augustine says in conclusion- was therefore not sufficient to bring about the fall of humanity.

The purpose of this chapter has been to raise questions and

issues by close scrutiny of the text of the *de Libero Arbitrio*. In subsequent chapters these issues and questions will be addressed more thematically, and in relation to a wider selection of Augustine's writings. But as a sufficient idea has now been given of the kind of issues with which this thesis will have to concern itself in addressing Augustine's theory of the voluntary, my survey of the themes of the *de Libero Arbitrio* is now complete.

Chapter 6

Voluntariness and Responsibility in Augustine

1. Introduction
2. The Linking of Voluntariness and Responsibility
3. Two of Augustine's Conditions of Voluntary Action
4. 'Ignorantia' and 'Difficultas'
5. From the Earlier to the Later Theory
6. Conclusion

'We in remorse are a radical minority within the social work community. We believe that not every wrong in our society is the result of complex factors such as poor early learning environment and resultative dissocialised communication. Some wrong is the result of *badness*. We believe that some people act like jerks, and that when dealing with jerks one doesn't waste too much time on sympathy. They're jerks. They do bad things. They should feel sorry for what they did and stop doing it.'

(Garrison Keillor, 1989)

1. Introduction

My analysis of Augustine's doctrine of human freedom to act begins in the same way as my analysis of Aristotle's: with an

account of the negative conditions which Augustine thinks necessary for voluntary action. The same beginning, for the same reason: because the philosophical 'problem of freedom' is no more and no less than a problem in the theory of action. To be able to give the full conditions of *voluntary* action is to be able to give those of *free* action also.

First I argue (§2) that Augustine's interest in voluntariness stems from his concern, as an apologist, with responsibility for evil. Then (§3) I survey (some of) his conditions for voluntary action. It will appear that, like Aristotle's, Augustine's theory of the voluntary has an important negative aspect. For him too the voluntary is at least partly defined by exclusion. (Does his theory, again like Aristotle's, also have a positive aspect? I address this question in Ch.7.)

As I note in §4, fallen humans' actions, being vitiated by what Augustine calls 'ignorantia' and 'difficultas', fail to satisfy his own negative conditions for voluntary and responsible action. Nonetheless such actions are, apparently, responsible. I explain this (§5) by noting Augustine's crucial distinction between culpable and non-culpable ignorance and compulsion.

2. The Linking of Voluntariness and Responsibility

First, why should a Christian pastor and apologist be interested in an arcane philosophical question like the nature of

the voluntary? Augustine himself tells us why. In his youth he was much troubled by the evergreen question 'Unde malum?' (Conf.3.7, dLA 1.4). This question led him into Manichaeism. It was only when he had a satisfactory answer to it, in Christian Platonist terms, that Augustine returned to the Catholic fold.

The problem which Augustine referred to by his question 'Unde malum?' is this one:

'Credimus autem ex uno Deo omnia esse quae sunt; et tamen non esse peccatorum auctorem Deum. Movet autem animum, si peccata ex iis animabus sunt quas Deus creavit, illae autem animae ex Deo, quomodo non parvo intervallo peccata referantur in Deum.' (dLA 1.4)

Nietzsche's barbs, in my first epigraph to chapter 5, are aimed at Augustine. For it was Augustine's efforts to solve *this* problem, the problem of evil, that induced him to develop his 'free will defence', his version of the Biblical argument (Gen.3, Mk.7.14-23, Jas.1.13-15) that evil results from the exercise of creatures' God-given autonomy, not from God's own action: 'Quisque malus sui malefacti auctor est' (dLA 1.1). Hence Augustine, and his commentators, are usually quicker to spot the difficulties he makes in theodicy in the dLA than those he makes in the theory of action. This, of course, is for the good reason that theodicy, not theory of action, is what he thinks he is up to.

Consider the first argument in the dLA:

'Si Deum iustum fatemur, nam et hoc negare sacrilegium est, ut bonis praemia, ita supplicia malis tribuit... Si nemo iniuste poenas luit, quod necesse est credamus, quandoquidem divina providentia hoc universum regi credimus, illius primi generis malorum nullo modo, huius autem secundi auctor est Deus... Malefacta iustitia Dei vindicari. Non enim iuste vindicarentur, nisi fierent voluntate.' (dLA 1.1; cp. dVR 27)

The argument is this:

1. God punishes some actions.
2. God is just.
3. God's punishment would not be just unless the actions punished were sins.
4. Therefore some actions are sins. (1,2,3)
5. But no action is a sin unless it is voluntary.
6. Therefore some actions are voluntary. (4,5)

The argument shows clearly enough why Augustine is interested in voluntary action: because he sees a necessary connection between voluntariness and responsibility. If there are no creaturely voluntary actions, then the only voluntary agent in existence will be God. But if God is the only *voluntary* agent, then (Augustine thinks) he is, necessarily, also the only *responsible* agent. Everything, good or bad, results directly from divine agency. Therefore God himself is directly responsible for

everything, including all the evil in the world. Augustine thinks this an impossible position for a Christian. Hence, as he argues repeatedly throughout his career, Christians are committed to believing that God is *not* the only voluntary agent.

However, we must distinguish two claims about the relation of voluntariness and responsibility:

1. Any piece of behaviour is a responsible action if and only if it is a voluntary action.
2. Any piece of behaviour is a responsible action if and only if it is *either* (i) a voluntary action *or* (ii) a relevantly connected consequence, e.g. a foreseeable causal consequence, of a voluntary action.

When Augustine writes, e.g., that 'usque adeo peccatum voluntarium est malum, ut nullo modo sit peccatum, si non sit voluntarium' (dVR 27), we might think that he was arguing for (1). But (1) is an unsophisticated and implausibly strong thesis. We have already seen that Aristotle rejected it (NE 1114a4-13): the person who kills someone while blind drunk is responsible for that killing, not so much because he chose to kill them, as because he chose to get blind drunk. This distinction between (as we might say) direct and indirect responsibility is also defended by Augustine, both in his early and in his late works. However, the importance of that distinction, as supporting a more fully developed doctrine of original sin, is clearer in his later work:

'Dereliquit [homo] Deum et factus est malo dignus aeterno, qui hoc in se peremit bonum, quod esse posset aeternum. Hinc est universa generis humani massa damnata; quoniam, qui hoc primus admisit, cum ea quae in illo fuerat radicata sua stirpe punitus est, ut nullus ab hoc iusto debitoque supplicio nisi misericordii et indebita gratia liberetur.' (dCD 21.12)

So Augustine's 'free will defence' is never simply (1), that the human agent is responsible directly for each and every wrongdoing she commits, because each such wrongdoing is itself a voluntary action. It is, more subtly, (2): that one original wrongdoing by a human agent was a voluntary and hence responsible action, and that all human wrongdoings since have been relevantly connected (how?) consequences of that responsible action.

3. Two of Augustine's Conditions of Voluntary Action

It is an important claim to say that voluntariness and responsibility are necessarily connected. Augustine never abandons this claim, at least not in the qualified form (2). But the claim is not, in itself, very informative. The prior and more important question is: What *is* 'voluntariness'? That is: Under what conditions is action properly called voluntary?

Augustine's responses to that question tend, as the dLA illustrates, to be unsystematic. Hence it is that, particularly

'Dereliquit [homo] Deum et factus est malo dignus aeterno, qui hoc in se peremit bonum, quod esse posset aeternum. Hinc est universa generis humani massa damnata; quoniam, qui hoc primus admisit, cum ea quae in illo fuerat radicata sua stirpe punitus est, ut nullus ab hoc iusto debitoque supplicio nisi misericordii et indebita gratia liberetur.' (dCD 21.12)

So Augustine's 'free will defence' is never simply (1), that the human agent is responsible directly for each and every wrongdoing she commits, because each such wrongdoing is itself a voluntary action. It is, more subtly, (2): that one original wrongdoing by a human agent was a voluntary and hence responsible action, and that all human wrongdoings since have been relevantly connected (how?) consequences of that responsible action.

3. Two of Augustine's Conditions of Voluntary Action

It is an important claim to say that voluntariness and responsibility are necessarily connected. Augustine never abandons this claim, at least not in the qualified form (2). But the claim is not, in itself, very informative. The prior and more important question is: What *is* 'voluntariness'? That is: Under what conditions is action properly called voluntary?

Augustine's responses to that question tend, as the dLA illustrates, to be unsystematic. Hence it is that, particularly

in Augustine's later work, crucial ambiguities and contradictions on that question pass (conveniently?) unnoticed. Augustine's attention, as we have seen, is on theodicy, not on action theory. Anyway, even the young Augustine usually sees no need to spell out the conditions of voluntary action because he thinks that they are intuitively obvious. Typically, his remarks about them are flanked by bold phrases like 'quis dubitet...?' (dDA 1.15), 'cum... natura ipsa proclamet' (dDA 1.14), 'Nonne ista cantant et in montibus pastores, et in theatris poetae, et indocti in circulis, et docti in bibliothecis?' (dDA 15). At dVR 27 we have an appeal to the authority of common sense reminiscent of Aristotle's appeals to 'the many and the wise' (NE 1095a18: v. Ch.2, §2):

'Hoc quidem ita manifestum est, ut nulla hinc doctorum paucitas, nulla indoctorum turba dissentiat.' (dVR 27)

This tone of airy confidence might encourage an unwary commentator to take Augustine's conditions of voluntary action for granted too. But a clear account of these is of the greatest importance for an adequate understanding of the strengths and weaknesses of Augustine's position. And, in spite of Augustine's tendency to gloss over them or only mention them in passing, it is not impossible to give an account of what these conditions are, as represented in the early works. I will base this account on the dLA (Augustine's fullest treatment of the issues) and the dDA (his ablest treatment).

The picture that emerges is remarkably Aristotelian (cp. Ch.1). For a start, the early Augustine agrees with Aristotle in believing that any behaviour, to qualify as voluntary action, must (at the least) satisfy two negative conditions, with which I will deal in this chapter. It must be:

- (i) not compelled,
- (ii) not done in ignorance.

(i) *Compulsion*.— dDA 14 gives us one of Augustine's most adequate formulations of the definition of voluntary action: 'Voluntas est animi motus, cogente nullo, ad aliquid vel non amittendum, vel adipiscendum'.

(I call this a definition of *voluntary action*. It may rightly be objected that, strictly, what Augustine is defining at dDA 14 is not *voluntary action* but '*voluntas*' (however we are to translate that). However, not much hangs on this here; see Ch.7, §2, for further discussion of the point.)

This definition, defended phrase by phrase in the dDA, covers nearly everything that the young Augustine believes about voluntary action; except, surprisingly, the question of ignorance. The role of the phrases 'animi motus' and 'ad aliquid vel non amittendum, vel adipiscendum' in Augustine's *positive* theory of the voluntary will be discussed below in §6. The words 'cogente nullo' cover the question of compulsion. Why does Augustine think these words should be included in the definition

of voluntary action? He writes elsewhere (dLA 3.51):

'De Dei iustitia dubitare dementis est... non enim quisquam iniustus dominator aut surripere hominem potuit... aut extorquere invito, tanquam invalidiori, vel terrendo vel conflagando, ut hominem iniusta poena cruciaret.'

And in the dDA:

'Omnis qui volens facit, non cogitur; et omnis qui non cogitur, aut volens facit, aut non facit.' (dDA 14)

Augustine thinks that his claim that voluntary action must be uncompelled action is evident, not just from Scripture, but even from nature:

'Haec cum in omnibus hominibus, quos interrogare non absurde possumus, a puero usque ad senem, a ludo litterario usque ad solium sapientis, natura ipsa proclamet...' (dDA 14)

In the dLA, Augustine relies repeatedly- though not always explicitly- on the premiss that wrongdoing, to count as sin, must be uncompelled. I give three examples.

(a) One point of Augustine's long argument (dLA 1.16-1.22), that nothing is superior in power to the virtuous soul, except Reason and God, is that it helps him to establish that the virtuous soul cannot be compelled to do evil. For if it was

compelled it would not be acting voluntarily, and so not sinning.

'Ergo relinquitur ut quoniam regnanti menti compositique virtutis, quidquid par aut praelatum est, non eam facit servam libidinis propter iustitiam... nulla res alia mentem cupiditatis comitem faciat, quam propria voluntas et liberum arbitrium.' (dLA 1.21)

(b) In Ch.5 I noted that Rist must be wrong to think that, for Augustine, there are no internal compulsions, and adduced Augustine's contrast between the stone and the actions of the human agent in the dLA. (Cp. Aristotle, NE 1103a18-20: δῆλον ὅτι οὐδεμία τῶν ἠθικῶν ἀρετῶν φύσει ἡμῖν ἐγγίνεται.) As I said there: If the agent were compelled as a stone is compelled, then of course the resulting behaviour would not be voluntary action or relevantly connected behaviour, and hence not directly or indirectly responsible action, and hence not sin. 'Quaecumque ista causa est voluntatis, si non ei potest resisti, sine peccato ei ceditur: si autem potest, non ei cedatur, et non peccabitur.' (dLA 3.50)

(c) In dLA 3.4-11, Augustine argues that his doctrine of God's foreknowledge of human actions does not endanger his theory of human freedom; what God foreknows is that the human agent will act *freely* (dLA 3.8). Foreknowledge presents the threat that God's foreknowledge be seen to *compel* our actions; and this would mean that they were not voluntary actions, nor relevantly connected behaviour. But 'Deus neminem ad peccandum cogens

praevideat tamen eos qui propria voluntate peccabunt' (dLA 3.10).
So 'Cur ergo non vindicet iustus, quae fieri non cogit
praescius?' (3.11).

To return to the dDA 14 passage with which I began my
consideration of compulsion. The claim made there by Augustine
is, apparently, that 'X acts voluntarily' entails 'X is not
compelled'. This claim will be true if the absence of compulsion
is either necessary, or necessary and sufficient, for voluntary
action. Obviously, since I say that the absence of compulsion is
not the only condition of voluntary action, my account requires
that this condition should only be necessary. But Augustine's
'Everyone who is not forced, either acts voluntarily or does not
act at all' is, apparently, a claim that the absence of
compulsion is both necessary *and* sufficient for voluntary action.
So where does this leave the negative condition of *ignorance*?

(ii) *Ignorance*.— In spite of this passage, there is plentiful
evidence that Augustine did not, in fact, normally neglect the
idea that ignorance is a limiting factor on responsible action.
Elsewhere in his early writings Augustine clearly does defend
that idea, so that the remark of dDA 14 is probably best
dismissed as a rhetorical excess, and not representative of
Augustine's usual theory of responsibility. I give, again, three
examples of this defence.

(a) In the dDA itself, at Ch.12, Augustine considers the
example of a person whose hand is made to write something obscene

while he is asleep. His being asleep exculpates him from the sin of writing rude words either (i) if he is compelled (as above), or else (ii) if he does not *know* that his hand is going to be used in this way while he sleeps. But if he does know, and does nothing about it when he could have, then 'Num ei quidquam somnus ad innocentiam suffragaretur?' (dDA 12). Strictly speaking, he will then be indirectly responsible, like the drunken killer. His knowledge meant that he could have avoided getting in a state where his hand would write filth, but he didn't.

(b) In the early RomsIE, Augustine discusses the 'unforgivable' sin of speaking against the Holy Spirit (Mt.12.32). He argues against the rigorous (Novationist) view that those who so sin after baptism cannot be forgiven, whereas those who speak against the Holy Spirit before baptism can be. The Novationists said that someone who speaks against the Holy Spirit after baptism could no longer be excused on the ground of ignorance. Augustine counters that it is perfectly plain that people can and do remain ignorant of such important teachings after baptism:

'Quid autem de his qui cum baptismi sacramenta pueri vel etiam infantes perceperint, postea negligenter educati per ignorantiae tenebras vitam turpissimam ducentes nescientes omnino, quid christiana disciplina iubeat aut vetet? Num audebimus peccata eorum propterea non ignorantiae peccata deputare, quia baptizati peccaverunt?' (RomsIE 16)

What is at issue between the Novationists and Augustine is: How wide is the scope of exculpating ignorance? What is not at issue, but is accepted on both sides, is that there *is* exculpating ignorance. (Note, however, that here Augustine seems prepared to agree with the Novationists that 'sins of ignorance' are nonetheless *sins*. What he is arguing is that such sins are *less* culpable, not that they are not culpable at all.) Augustine was arguing that the Novationists' understanding of that notion was insufficiently generous.

(c) At dLA 3.50, Augustine considers the idea that the experience of deception can sometimes result in a form of ignorance which exculpates:

'An forte fallit incautum? Ergo caveat ne fallatur. *An tanta fallacia est, ut caveri omnino non possit? Si ita est, nulla peccata sunt.* Quis enim peccat in eo quod nullo modo caveri potest? Peccatur autem: caveri igitur potest.' (dLA 3.50)

So, despite the evident claim of dDA 14, the early Augustine *does* in fact believe that there can be exculpating ignorance, parallel in kind to exculpating compulsion.

4. 'Ignorantia' and 'Difficultas'

However, as early as the dLA there is evidence of the beginnings of a quite different attitude to ignorance and

compulsion, an attitude which came to dominate Augustine's later theory of responsibility. For the classical tradition, and with it the early Augustine, the conditions of ignorance and compulsion were *restrictions* on culpability. But in Augustine's later writings on the subject, it often seems as if what he calls the conditions of 'ignorantia' and 'difficultas' have become the very *hallmarks* of culpability.

In his later works, as I will now show, Augustine often seems prepared to argue that sin, so far from being necessarily—neither compelled nor done in ignorance, is *typically* one or the other, or both. Augustine never renounced any of the central points of his teaching on the subject of responsibility and voluntariness. Yet his emphasis changes so dramatically that it looks at first as if he has done a complete *volte face*: from the view that 'ignorantia' and 'difficultas' are conditions which diminish responsibility; *via* the position that, as 'penal conditions', they are characteristic of fallen humanity; to the view that these conditions are actually culpable in themselves.

'Nam sunt revera omni peccanti animae duo ista poenalia, ignorantia et difficultas' (dLA 3.52). As Augustine developed his theory of human nature, he came to think that fallen humans were not just thwarted from the achievement of their wishes by ignorance and compulsion in general. They were, rather, typically subject to *specific forms* of ignorance and compulsion. What were these forms?

In dLA 3.51-6 Augustine says that fallen man acts wickedly either because 'non habet in potestate ut bonus sit', or by 'non videndo qualis esse debeat' (3.51). It is already apparent that his 'difficultas' and 'ignorantia' are impediments to voluntary action of quite specific kinds. 'Difficultas' characterises the agent who wishes to do what is right, but cannot. 'Ignorantia' characterises the agent who does not even have the knowledge of the good that puts him in a position to wish to do what is right.

(Does this mean that Augustine's 'ignorantia' is equivalent to Aristotle's 'ignorance of principle' rather than Aristotle's 'ignorance of particulars'? It seems that Augustine does not really take this distinction on board. He is more concerned to make out his own distinction, between voluntary and involuntary ignorance, which (he thinks) is the most morally important distinction. The Aristotelian question, whether there is any difference in moral status between voluntary and involuntary ignorance of principle, apparently does not occur to him.)

Augustine gives us an even more detailed account of the roles of ignorance and difficulty in thwarting voluntary action in four works of the early 390s (EPRoms, Simp, 83DQ, EEGal). Here he sees human moral regeneration as proceeding through four stages, summarised by EPRoms 13:

'Ante legem sequimur concupiscentiam carnis, sub lege trahimur ab ea, sub gratia nec sequimur eam nec trahimur ab ea, in pace nulla est concupiscentia carnis'.

Ante legem there is complete 'ignorantia', but absolutely no 'difficultas'. There is no struggle (EPRoms 14, 'non pugnamus'), because the human is in a state of 'concupiscentia', in which she simply follows her natural inclinations, and even approves of so doing (EpRoms 14, 83DQ 66.3). There is nothing to prevent this, for the human has no conception that what she does is wrong (EEGal 46, 'non est qui prohibeat'). 83DQ 66.4 gives four Biblical proof texts: Roms.5.12-13, 7.8-9, 7.13, and 1 Tim.1.8.

Sub lege there is no important 'ignorantia', but plenty of 'difficultas'. The human experiences unsuccessful struggle (EPRoms 13, 'trahimur'; DQ83, 'victi peccamus') against desires which are now perceived, by the recognition which the Law brings, as evil (EPRoms 15 'Fatemur mala esse quae facimus'; EEGal 46 'Conatur a peccato abstinere se, sed vincitur'). EEGal adds that the struggle is unsuccessful because 'nondum iustitiam propter Deum et propter ipsam iustitiam diligit, sed eam sibi vult ad conquirendum terrena servire'. Paul's Manichaean sounding remark that his sins are due not to himself but to ἡ ἐνοικοῦσα ἐν ἐμοὶ ἀμαρτία (Roms.7.18, cp. *de Duabus Animabus*) is picked up by EPRoms 15: 'fatendo mala esse utique nolumus facere'; hence (says EEGal 46) the sinner 'trahitur pondere temporalis cupiditatis, et relinquit iustitiam'. This remark has a very Platonic ring (cp., e.g., *Republic* 439a-440d). As proof texts 83DQ 66.5 cites Roms.5.20, most of Roms.7.5-25, and Ps.18.13.

Sub gratia there is no 'ignorantia', and less and less 'difficultas', since the agent's experience is now of successful

struggle with sin (EEGal 46):

'In hac enim vita etiamsi existant desideria carnis de mortalitate corporis, tamen mentem ad consensionem peccati non subiugant.'

Cp. 83DQ 66.3:

'Iam non vincimur delectatione consuetudinis malae... sed tamen adhuc eam interpellantem patimur, quamvis ei non tradamur.'

This victory is, of course, the result of grace (EPRoms 16):

'Venit ergo gratia, quae donet peccata praeterita et conantem adiuvet et tribuat caritatem iustitiae'.

Without this grace victory is impossible (EEGal 46):

'Nisi charitate spirituali quam Dominus exemplo suo docuit et gratia donavit, fieri non potest.'

Augustine's proof texts are: Roms 7.25, 8.1-10, 25 (83DQ 66.6).

In pace there is no 'difficultas' and no 'ignorantia'. There is now no struggle with sin, for sin is finally defeated and the agent is perfected:

'postea vero ex omni parte exstinguitur.' (EEGal 46)

'Quarta est actio, cum omnino nihil est in homine quod resistat spiritui, sed omnia sibimet concorditer iuncta et connexa unum aliquid firma pace custodiunt.' (83DQ 66.3)

This stage, of course, is only reached in the life of the Resurrection:

'...Quod fiet mortali corpore vivificato.' (83DQ 66.3)

'Ideo autem perfecta pax, quia nihil nobis resistet non resistantibus Deo.' (EPRoms 17)

As his scriptural warrant for this Augustine cites Roms 8.11 (83DQ 66.7).

All four of these stages are implicitly identified at Simp 1.2:

'Legem ad hoc datam esse... ut peccatum demonstraretur, quo animam humanam quasi de innocentia secunda ipsa peccati demonstratione ream faceret: ut, quia peccatum sine gratia Dei vinci non posset, ipsa reatus sollicitudine ad percipiendam gratiam convertitur.'

First, the state of 'carefree innocence'; or rather carefree quasi-innocence, for of course Augustine does not mean that the

man 'ante legem' is sinless, just that her sins are unrevealed
Simp 1.6: 'Non quia [lex] non erat, sed quia non apparebat').
then the stage of a 'sollicitudo' caused by the awareness of the
law; then, less clearly here, the two stages of grace and peace.

(With this four-stage progress from wickedness to holiness,
compare four Aristotelian states: ἀκολασία, ἀκρασία, ἐκρατεία,
and σωφροσύνη.)

Simp 1.11 gives us an opinion on the interesting question of
whether 'difficultas' applies just to the *performance* of good
actions, or even to the *willing* of them. There Augustine notes,
of the 'sub lege' stage:

'Certe enim ipsum velle in potestate est, quoniam adiacet
nobis; sed perficere bonum non est in potestate.' (Simp 1.11)

'Sine difficultate vult, quamvis non tam facile faciet, quam
facile vult.' (Simp 1.12)

Perhaps we may say that, at the first stage ('ante legem'),
what the agent lacks, being in a state of 'ignorantia', is both
good intentions and good performances. Whereas at the second
stage ('sub lege'), the agent has good *intentions*, but he is
(through 'difficultas') incapable of any corresponding (good)
performances. The role of grace, at the third and fourth stages,
is to connect willing with performance. It is to be noted that in
this area Augustine's doctrine does appear to have changed

radically. The later Augustine quotes repeatedly the dictum of Paul that 'God is at work in you *both* to will and to work according to his good pleasure' (Phil.2.12-13).

It is now pretty clear what Augustine means when he talks of 'ignorantia' and 'difficultas'. The next question is: How did these conditions, so far from being thought exculpatory, come to be seen by Augustine as typical or even diagnostic of culpability?

5. From the Earlier to the Later Theory

What enabled Augustine to take this direction was his development, again from what was originally fairly uncontroversial classical material, of the distinction between non-culpable and culpable ignorance and compulsion. He came increasingly to stress the culpable and neglect the non-culpable, until by the late works we hear almost nothing of exculpatory ignorance and compulsion, and a very great deal indeed about culpable ignorance and compulsion.

But it should be noted that what was involved was more a change of emphasis than of doctrine. This is evident from the dLA, in which most of the doctrines usually supposed to characterise only Augustine's *late* theology can already be found. The dLA, it is true, teaches such typically 'early' doctrines as synergism, the view that God and human work *together* in the

economy of salvation. (So, e.g., when Augustine says that grace is given to those whom God foresees will respond to it: dLA 3.33.) But side by side with this kind of teaching, often in the same breath, comes the 'later' stress on grace. By and large, Augustine, in the course of his career as a writer on freedom and the voluntary, did not so much change his position as his nuances- promoting now one side of what he saw as Biblical doctrine, and now the other. This can be nicely illustrated by showing how, for Augustine's 'late' and 'early' doctrines on the issue of culpable and non-culpable ignorance and compulsion, we need look no further than the dLA, where nearly all of them can be found jumbled together.

If we may cut through the attritions, we may suggest that Augustine's most basic teaching on this issue is that 'natural' ignorance and compulsion are not culpable, but 'voluntary' ignorance and compulsion are. Natural ignorance and compulsion are just where the soul begins its journey, even in an unfallen creation. So, at dLA 3.71, Augustine has to deal with this dilemma:

'Si sapiens factus est primus homo, cur seductus est? Si autem stultus factus est, quomodo non est Deus auctor vitiorum?'

His response to this challenge is that:

'Est enim stultitia, rerum appetendarum et vitandarum non quaelibet, sed vitiosa ignorantia.' (dLA 3.71)

Which ignorance, then, is vicious (and culpable), and which innocent (and blameless)? Augustine's answer is that the innocent form of ignorance is that which is not chosen by the agent, viz. natural ignorance. *Pari passu*, the innocent form of compulsion is natural compulsion. Actions done under these influences are not culpable for Augustine because *you can't avoid them* (cp. dLA 3.50, quoted above). This was still Augustine's view when he wrote the dGLA: 'Aliud est enim nescisse, aliud scire noluisse' (dGLA 5).

This is a perfectly mainline view, in which no important ancient philosopher would have seen any problem. But what are 'voluntary' ignorance and compulsion, and why are they culpable in fallen humans? This is where Augustine's teaching in the dLA becomes confused. Sometimes (an 'early' doctrine, consistent with the view of 'natural' compulsion just outlined) he tells us that they are culpable in fallen humans because fallen humans can themselves avoid them:

'Non tibi deputatur ad culpam quod invitus ignoras, sed quod neglegis quaerere quod ignoras; neque illud quod vulnerata membra non colligis, sed quod volentem sanare contemnis.' (dLA 3.53)

'Non enim quod naturaliter nescit et naturaliter non potest, hoc animae deputatur in reatum; sed quod scire non studuit, et quod dignam facilitati comparandae ad recte faciendum operam non dedit.' (dLA 3.64)

At other times, however, we get the 'later' doctrine that they are culpable in fallen humans because Adam and Eve could have avoided them:

'Illud quod ignorans quisque non recte facit, et quod recte volens facere non potest, ideo dicuntur peccata, quia de peccato illo liberae voluntatis originem ducunt: illud enim praecedens meruit ista sequentia.' (dLA 3.54)

'Ut autem de illo primo coniugio, et cum ignorantia, et cum difficultate, et cum mortalitate nascamur, quoniam illi cum peccavissent, et in errorem, et in aerumnam, et in mortem praecipitati sunt, rerum moderatori summo Deo iustissime placuit.' (dLA 3.55)

At other times again it is already beginning to appear in the dLA as if such conditions are *culpable in themselves*:

'Si non est ista [sc. facta ex ignorantia et difficultate] poena hominis, sed natura, nulla ista peccata sunt.' (dLA 3.51)

The question raised by this apparent diversity of opinions is one of culpability. Is human responsibility for sin a matter of what, above, I called *direct* culpability, or of *indirect* culpability, or of a mixture? Which of these is Augustine's claim?

(a) There are wrong actions which fallen humans perform for which they are directly to blame.

(b) There are wrong actions which fallen humans perform for which they are directly to blame, and as a result of which they helplessly perform other wrong actions for which they are indirectly to blame.

(c) There are wrong actions which fallen humans perform for which they are indirectly to blame.

When Augustine talks of ignorance and compulsion as penal conditions, as he often does even in the dLA, it seems at first as if he is defending (a) above. So, for example, 3.52: 'Illa est enim peccati poena iustissima, ut amittat quisque quo bene uti noluit'. The person who fails to take the initiative for good in her own life is punished by the removal of the chance to take that initiative at all.

But then we notice that actions performed in 'ignorantia' and 'difficultas' are themselves, sometimes, punishable:

'Etiam per ignorantiam facta quaedam improbantur... sunt etiam necessitate facta improbanda, ubi vult homo recte facere, et non potest; nam unde sunt illae voces, "Non enim quod volo facio bonum, sed quod nolo malum, hoc ago"?' (dLA 3.51).

The idea (b) that the kind of culpability in question for such actions is mainly indirect might seem to be supported by Augustine's contention, in at least two other places, that sins

of ignorance are culpable, but *culpable in a reduced degree*. So in RomsIE 17 Augustine entertains the conclusion that the only venial sins are those committed in ignorance: all others, he seems to be arguing here, are mortal. In dGLA 5, Augustine quotes the same Lucan verses (Lk.12.47-48) in support of a somewhat similar argument, that 'gravius peccare hominem scientem quam nescientem'.

But this argument too seems to be denied by the assertion above, from dLA 3.54, that *all* fallen human sinfulness is indirect and dependent on the wrong choice of Adam and Eve. So Augustine's position becomes still harder to elucidate: it now seems closer to (c).

In any case there is a logical problem with (c), inasmuch as indirect responsibility is 'parasitic', logically dependent, on direct responsibility. At least in the sense in which 'responsible' is normally understood, I cannot be indirectly responsible for any doing of mine B unless (i) I was directly responsible for some other doing of mine A, and unless (ii) B was a relevantly connected consequence, e.g. a foreseeable causal consequence, of A. But Augustine's doctrine of responsibility seems to fall at both these hurdles. (i) If my culpability under the doctrine of original sin depends on what *Adam and Eve* did, then it does not depend on any doing of mine, responsible or otherwise. (ii) *A fortiori*, the consequences of what they did were not, and could not have been, 'relevantly connected' to their action(s) *from my point of view*, neither as 'foreseeable

causal consequences' nor as anything else.

Augustine seems aware of this problem, and it leads his theory into one last contortion, not yet noted, but hinted at in one of the quotes above:

'Si non est ista [sc. facta ex ignorantia et difficultate] poena hominis, sed natura, nulla ista peccata sunt.' (dLA 3.51)

Augustine's last manoeuvre on this topic, only performed on the occasions when he thinks he can get away with it, is actually to claim that what we are punished for by God is *what we do in ignorance and difficulty*. The penal conditions themselves become *their own punishment*. To offer two examples of this rare, spectacular back-flip:

(a) dCD 14.11: Eve was guilty *because* she was 'seduced', i.e. deceived (and therefore ignorant). And Adam, says Augustine, 'was not *less* (!!) guilty because he was *not* deceived'.

(b) dGLA 5: 'When a man says, "I cannot do what I am commanded, because I am mastered by my own concupiscence", he has no longer any reason to blame God in his heart, but he recognises and laments *his own evil in himself*.'

If one is attempting, as Augustine is, to explain the performance of large numbers of evil deeds by humans, this last manoeuvre is, of course, a patently circular and self defeating one; not so much a back flip as a belly flop. One might as well

'explain' the presence of large numbers of prisoners in jails by arguing (legitimately) that they were there because they were being punished, and then (crazily) that what they were being punished for was the offence of being prisoners.

56. Conclusion

I have argued in this chapter that Augustine quite often articulates a very standard classical view of ignorance and compulsion as exculpating or mitigating factors, and that nothing he ever says amounts to a serious denial of this standard view. But, as I have also pointed out, in certain moods Augustine can be seen as attempting, in effect, to stand the classical view on its head. For sometimes he argues that certain kinds of ignorance and compulsion not only are not exculpating factors: they are either standard features of the behaviour of agents who typically incur blame; or else standard features of blameworthy cases of action; or even good reasons for ascribing blame to actions in which they are evident.

This attempted reversal is a bold and interesting move, but, as Augustine presents it, it does not work. I have pointed out in his work some of the various theses about responsibility and voluntariness which Augustine sometimes seems prepared to endorse. But it is not satisfactorily clear exactly which of these he means, overall, to be arguing; and he cannot coherently combine them. Augustine starts from the tried and tested

foundation of eclectic classicism, and launches zealously out on a highly revisionary programme in the philosophy of action. This heroic programme might have been more successful had Augustine paid closer and more sustained attention to those fine logical details and distinctions which, at his best, he himself delighted in drawing.

Chapter 7

'Voluntas' and the Voluntary

1. Introduction
2. 'Voluntas': 'Voluntary Action' or 'Volition'?
3. The Nature of 'Voluntas': Two Requirements
4. Ability to Do Otherwise
5. The Cause(s) of 'Voluntas' again
6. A Reflexive 'Voluntas'?

'Si liberum non sit, non est voluntas.'

(Augustine, *de Duabus Animabus*, 15)

1. Introduction

In the last chapter, I examined two negative conditions which Augustine gives for voluntary action, namely that such action should (i) not be compelled and (ii) not be done in ignorance. I also raised the question whether Augustine, like Aristotle, gives any positive conditions for voluntary action.

In this chapter and the next, I argue that Augustine does indeed give us at least two such conditions of normal voluntary action, (i) and (ii):

(i) Voluntary action, for Augustine, is action on that 'animi motus' which he calls the 'voluntas', and as such is the very opposite of compelled behaviour.

(ii) Augustinian voluntary action is also action with the form of an aspiration to the good life ('ad aliquid vel non amittendum vel adipiscendum'). As that good life is, in Augustine's view, of determinate form, such aspirations partake of the 'measure, weight and number' of practical wisdom; which makes them anything but ignorant.

(i) is a converse of the negative condition about compulsion. (ii) is, as I will argue, a positive account of the *rationality* of Augustinian voluntary action. Its existence suggests the existence also of a negative condition of voluntary action, that it should not be irrational; but no separate attempt will be made to draw this out. Nor will I discuss Augustine's theory of knowledge, the positive converse of his negative condition about ignorance. No doubt such a converse can be found in Augustine's Platonist epistemology; but this subject is too large to discuss in a thesis on his theory of freedom, and (unlike Aristotle's theory of knowledge) is in any case not neatly separable from what Augustine has to say about practical rationality/ wisdom.

In this chapter I will deal with (i), Augustine's theory of the 'voluntas'- his account of the efficient causation or origination of voluntary action. In Ch.8 I will deal with (ii), Augustine's theory of 'felicitas'- his account of such action's

final and formal causation.

2. 'Voluntas': 'Voluntary Action' or 'Volition'?

What, in Ch.6, I called the negative conditions of voluntary action, Augustine discusses as constraints on '*voluntas*'. Hence I was content there to treat Augustine's '*voluntas*' as if it meant simply 'voluntary action'. But does '*voluntas*' mean 'voluntary action' or the *cause* of such action, *i.e.* 'volition'?

Quite often, Augustine uses '*voluntas*' to distinguish, not volitions from actions, but *voluntary actions* from other behaviour. When, e.g., he writes (dLA 1.30) that humans come to merit the good life '*voluntate*', he does not mean that they merit it simply by choosing well, but by both choosing *and* doing well. Likewise, those condemned for '*mala voluntas*' are not being condemned for evil *volitions*, but for evil *voluntary actions*. Again, when (dDA 15) Augustine defines sin as '*voluntas retinendi vel consequendi quod iustitia vetat, et unde liberum est abstinere*', it seems that this use of '*voluntas*' admits of either translation.

But Augustine also uses '*voluntas*' to mean 'volition', and not 'voluntary action':

'Quid est enim quod facit voluntatem malam, cum *ipsa faciat opus malum*? Ac per hoc *mala voluntas efficiens est operis mali...*' (dCD 12.6)

'Non enim quidquam tam firme atque intime sentio quam me habere voluntatem, *eaque me moveri ad aliquid fruendum.*' (dLA 3.3)

By extension from Augustine's use of 'voluntas' as 'volition', there has arisen the common view, most recently defended in Dihle (1982), that his whole moral theory is based exclusively on 'voluntas' as volition. For him (it is suggested) the moral nature of an action is determined solely by the 'voluntas' from which it results. The whole of our moral theory, including the assessment of responsibility, can be concentrated on the volitions, and may treat the resulting actions as peripheral to the main issue. 'Interest *qualis sit voluntas hominis*' (dCD 12.6).

There is no doubt that Augustine, like his commentators, is influenced by this kind of thinking, traceable to Jesus' remarks on the origins of human uncleanness in the *heart* (Mt.15.22). However, it is simply untrue that Augustine places all moral significance in the volition (as opposed to voluntary action). First, as we have just seen, 'voluntas' does not only mean 'volition'. Second, it would in any case have been implausible to place all moral significance in the volition. Clearly there *are* morally significant differences between wanting to kill someone and actually killing them, even if both are reprehensible. Augustine was as well aware of these differences as Christ was: compare Jesus' 'The spirit is willing but the flesh is weak' (Mt.26.41) with Augustine's remarks (Ch.6, 54) on 'difficultas'

as the difference between wanting to do good and actually doing it (e.g. 'In nobis autem ante gratiam non est liberum arbitrium ut non peccemus, sed tantum ut peccare nolimus' (EPRoms 18)).

Augustine does not have a well elaborated theory of action, like Aristotle's- and in any case, as we have seen (Ch.3), even Aristotle's theory displays an important ambiguity at an analogous point, about whether deliberation causes προαίρεσις or action on προαίρεσις. 'Voluntas' does not simply and exclusively mean either 'volition' or 'voluntary action'; it is a rather loose term which Augustine allows himself to use to mean both. Hence I will not translate it exclusively as either 'volition' or 'voluntary action'; and often I will mark the ambiguity by leaving it in the Latin.

(I should also point out that there are, in both cases, type/token ambiguities: does 'voluntas' mean 'volition (in general)' or 'a (single) volition'; 'voluntary action (in general)' or 'a (single) voluntary action'? Latin's lack of an indefinite article does not help here.)

3. The Nature of 'Voluntas': Two Requirements

What, then, does Augustine mean by talking of 'voluntas' as the origin, or alternatively the form, of voluntary action? I think this requirement breaks down into two separate subparts.

behaviour, to count as voluntary action, must be positively describable as being:

- (a) original to the agent in question; and
- (b) such that the agent could have chosen to do otherwise.

These conditions might seem to be equivalent. It might be said that any action which is original to me (a), is one I am in control of; and any action I am in control of, is one I can either do or not do— a description which satisfies (b). But, in fact, (a) and (b) are quite separate conditions. To see why, consider these two triplets of claims:

1. A chooses to do X; A does X; A could have done Y, if A could have chosen to do Y, and had chosen to do Y.
2. A chooses to do X; A does X; A could have chosen to do Y.

(In (1), 'A could have done Y, if A could have chosen to do Y, and had chosen to do Y' is a more formal representation of the claim we make by the loose locution 'He could have done it if he had chosen to'. It might seem that an adequate formalisation of this would be: 'A could have done Y, if A had chosen to do Y'. But this leaves as an implication the vital point about (1): namely, that choosing entails ability to choose.)

If (a) were equivalent to (b), (1) would entail (2); which it does not, because of the third part of each triplet. 'A could have done Y, if A could have chosen to do Y, and had chosen to do

y' does not entail 'A *could* have chosen to do Y' (and/ or: '...A *did* choose to do Y'). The former does not entail the latter: any more than 'If Brownies were little boys, they would be Cubs' (which seems to be true) entails 'Brownies *are* little boys' (which is certainly false). The move from the conditional to the actual is illegitimate. Therefore (a) and (b) are not equivalent.

To specify some item A as the *origin* of a 'motus' M, is intended to mean that A was the *cause* of M: without A, M would not have occurred. Now in the case of a voluntary action, we say that the efficient cause of the 'motus' M (the behaviour, not the action) is the *agent* A, or (more precisely) A's decision to act and her execution of that decision. So M would not have occurred without A's decision to act and her execution of that decision. This in turn implies that, had A decided otherwise, a different 'motus', or no 'motus' at all, would have been the outcome.

This is to say that A's production of M was such that A *could* have done otherwise, had she been able to choose to do otherwise, and had so chosen. But- and this is the vital distinction- it is not to say that A *could* have chosen to do otherwise.

Causal statements (I take it) depend for their truth on the truth of *whole* conditionals. But the truth of a whole conditional statement need not depend on the truth either of its protasis or of its apodosis. For it to be true that q caused p, it only needs to be the case that *if not-q*, then not-p. It does not (ever) need to be the case that *not-q*. Likewise, for it to be true that the

cause of M was the agent A's decision to act and her execution of that decision, it only needs to be the case that *if* A had not taken that decision and acted on it, M would not have occurred. It does not (ever) need to be the case that A *could* have taken, let alone *did* take, a contrary decision.

Hence any behaviour M (it might be said) is coherently characterisable as a 'voluntary' action in two quite different senses. First, because M was caused by some agent A: i.e., M would not have occurred without A's decision to act and her execution of that decision. This weaker sense is coordinate with condition (a). And second, because- in addition to what the first sense specifies- M is also something that A could have chosen to do *and* could have chosen not to do: a sense in line with both conditions.

Condition (a) might be made the sole positive condition of voluntariness, yielding a thin account of voluntary action which is compatible with strict determinism. Or conditions (a) and (b) could be combined, to give a rich account of voluntary action, one which makes room (as the thin account seems not to) for a plausible theory of responsibility. I am arguing that Augustine's wider purpose of theodicy (among other considerations) commits him to argue, rightly, for the rich account. And he acknowledges that he is so committed; most often in his earlier writings, but not just there.

Nonetheless, the later Augustine seems, at times, to have

been attracted by the thin account. In that mood he will, as we have seen, come close to arguing that one's inability to choose to do otherwise than sin is no reason for remitting one's culpability- and indeed may actually be *evidence* of one's culpability. But he himself presented good reasons for thinking this position untenable; as we shall see, when we look at the evidence for Augustine's acceptance of (a) and (b).

(a) As we have seen, the condition that I should be the origin of those actions which truly pertain to me, simply means that it should be *my* 'animi motus', *my* 'voluntates' (in the sense of 'volitions'), which those actions, non-accidentally, express. Now there are, for Augustine, certain limitations, in the nature of things, on what shape my 'voluntates' could have.

Two in particular: first, my tendency to have 'bonae' or 'males' 'voluntates' is governed by my *habitation*, my 'consuetudo' as Augustine calls it (v., e.g., 83DQ 40). Second, a direct desire for something which I hold to be evil, desired *because* it is evil, could never be a 'voluntas' of any agent created by a good God. In Augustine's view, all 'voluntates', good or bad, aim at a single and particular good (however indirectly or inaccurately they may be aimed). However, to explore this side of the doctrine of 'voluntas' is really the topic of Ch.8; for this aspect of the doctrine has more to do with Augustine's teachings about practical rationality and wisdom than with his teachings about 'facilitas' or 'libertas' (as the converse of 'difficultas' or 'violentia').

Although, as I have stressed, (a) and (b) are logically separate conditions, still Augustine would insist on their contingent connection in human psychology. For him, a different way of defining what it means to require that voluntary action should be original to the agent, would be to say that to require this is to require that voluntary action should originate in that 'part of the agent' in which it could have originated otherwise. This leads us on to (b).

4. Ability to do Otherwise

That Augustine does take (b), ability to do otherwise, to be a condition of voluntary action is clear in at least four separate ways. Namely:

1. the connection between this condition and another already granted;
2. the historical context of his choice of the phrase 'liberum arbitrium';
3. his definition of 'peccatum'; and
4. his use of the language of the neutral will, or at any rate the morally ambivalent will.

1. As noted, positive condition (b) is connected to the negative condition that voluntary action should not be compelled (v. Ch.6). For, as Augustine points out, if it is not open to an agent to do otherwise than she actually does, then we usually

call her a compelled agent, not a voluntary one.

But, of course, this might be disputed. Is it self contradictory to claim both that 'A does X voluntarily at t', and that 'A has no choice but to do X at t'? It might seem not. To begin with, it may be said that 'has no choice' is an ambiguous phrase. A politician convicted of corruption 'has no choice' but to resign, and a sheep which falls out of a tree 'has no choice' but to plummet to the ground; but still we say that the politician resigns voluntarily and the sheep plummets involuntarily. So, it might be argued (say, by the kind of determinism sketched above), there are different senses in which A may have no choice but to do X at t, and not all of these are relevant to the question of voluntariness. In particular (it might be said), there is an important contrast between these situations:

- (1) A has no choice but to do X at t because A is not acting voluntarily; that is, A's doing X at t is not 'original to A' in the required sense.
- (2) A has no choice but to do X at t because A, although acting voluntarily at t, *never* has any choice to do or not to do.

To refer to the distinction made above: on this view, 'A does X voluntarily at t' entails only that A could have done other than X at t, had she been able to choose to do otherwise, and had so chosen. It does not entail that A *could* have chosen to do other than X at t.

For Augustine, the main problem with this position would be, as I have suggested, its incompatibility with any convincing account of responsibility. If we are to keep responsibility and voluntariness as tightly connected as (I argued in Ch.6) Augustine would like to, then what are we to say here? We cannot say that my responsibility, like my voluntariness, turns not on my ability to choose to do other than I actually do choose to do, but on its being a counterfactual truth that I could have done other than I did, had I been able to choose to do otherwise, and had I so chosen. For this formula (as shown) leaves the possibility open that I was not able to choose to do otherwise. But consider a case where I was *not*, in fact, able to choose to do otherwise than I did. It does not sound plausible to say that, in such a case, the right way to settle whether or not my action was responsible would not be to ask about *whether* I was unable to choose to do otherwise than I did, but about *why* I was unable to choose to do otherwise than I did.

2. The very phrase 'liberum arbitrium' implies this condition, as is shown by the *Oxford Latin Dictionary's* entry on the phrase (s.v. 'arbitrium').

The phrase has legal origins: the OLD cites Livy's use of it at *Historiae* 32.37.5. There Livy is describing a meeting between the Roman general Quinctius and Philip of Macedon. He writes: 'Sic infecta pace regii dimissi: Quinctio *liberum arbitrium* pacis et belli permissum' ('So the royal ambassadors were dismissed without obtaining peace: Quinctius was given *full discretion*

regarding peace and war'). Livy's point is that it was entirely up to Quinctius to declare or not declare war. No Roman law or regulation bound him to do either, and so the decision was deemed to rest with him. His action was such that he could have done otherwise had he chosen to.

Likewise, when the jurist Gaius wants to say that the court is not bound by any established rule to follow a particular procedure against a defendant, he says that the court '*liberum arbitrium* habet vel capitali crimine reum facere eum vel damnum persequi' (Gaius, *Institutiones*, 3.213). The point is exactly that the court is free to do either.

This historical evidence gives us good reason to presume that Augustine's choice of the phrase '*liberum arbitrium*' was partly motivated by its legal sense of 'freedom of judgement' or 'discretion'. Augustine deliberately used a term which implied, to Roman ears, that the kind of decision in question was the kind which can go *either way*.

3. At dDA 15, Augustine defines 'peccatum' thus:

'Peccatum est voluntas retinendi vel consequendi quod iustitia vetat, *et unde liberum est abstinere.*'

It will be objected that the italicised words give a constraint on what counts as a sin, not on what counts as a voluntary action. How can this remark help establish that (b) is

a condition of voluntary action? Doesn't it, rather, imply that there might be 'voluntates' from which one was *not* free to abstain? These objections are met when Augustine continues:

'*Quanquam si liberum non sit, non est voluntas... quod si nemo vituperatione vel damnatione dignus est, aut non contra vetitum iustitiae faciens, aut quod non potest non faciens, omne autem peccatum vel vituperandum est, vel damnandum; quis dubitet tunc esse peccatum, cum et velle iniustum est, et liberum nolle?*'
(dDA 15)

Here Augustine gives two necessary conditions for any deed to count as a sin: (i) it must be a wrongdoing, (ii) it must be avoidable. Thus an agent who does what it is impossible for him not to do, even if this is 'forbidden by justice', cannot be said to have sinned.

(N.B. the distinction, which will come up again in Ch.9, between 'wrongdoing' and 'sin'. A 'wrongdoing' is an infraction of the commandments of a legal or quasi-legal code. A 'sin' is such an infraction which is *deliberate*. Augustine's view on the relation between sin and wrongdoing switches back and forth between these four alternatives:

- (i) all sin (which is directly culpable) is wrongdoing, but not all wrongdoing (which may be either indirectly culpable, or not culpable at all) is sin;
- (ii) all wrongdoing is either directly culpable (sc., when it is

also sin) or indirectly culpable (sc., when it is merely wrongdoing);

(iii) the words 'sin' and 'wrongdoing' have exactly the same extensions: all wrongdoing is essentially sin, and there is no non-deliberate wrongdoing; and

(iv) the words 'sin' and 'wrongdoing' have exactly the same extensions: all sin is essentially wrongdoing, and there is no deliberate (in the sense of avoidable) wrongdoing at all; but we are nonetheless somehow to blame for our wrongdoings.

Augustine's best choice out of these four, and the position which he most coherently presents and should most consistently have presented, is (i). But it cannot be denied that, on *this* question, it is more than a shift of emphasis in his position that is evident when we compare the earlier and later works. V. Ch.6, §5.)

The action of such an agent as Augustine envisages at dDA 15 does not, according to (i), count as a sin because it does not count as a responsible action. It does not count as a responsible action because (v. Ch.6) it does not count as a voluntary action. And why does it not count as a voluntary action? Because- and this is the point which Augustine is making here- 'if the agent is not free [sc. to abstain], there is no voluntary action'. Hence the relevance of (b), the positive condition of avoidability, is not just to sins but to voluntary actions in general.

4. Augustine commonly seems to talk of 'voluntas' as being *neutral* between good and bad. I shall be arguing in Chs.8-9 that, on Augustine's conception, 'voluntas' is anything *but* neutral between good and bad (although it is, somehow, able to turn to either). But this argument might seem an uphill struggle when confronted with some passages from Augustine. Such as these three:

(i) 'Motus autem quo huc aut illuc voluntas convertitur, nisi esset voluntarius, et in nostra positus potestate, neque laudandus cum ad superiora, neque culpandus homo esset cum ad inferiora detorquet *quasi quendam cardinem voluntatis*.' (dLA 3.3)

(ii) 'Virtutibus nemo male utitur: ceteris autem bonis, id est, mediis et minimis, non solum bene, sed etiam male quisque uti potest... Voluntas ergo medium bonum est.' (dLA 2.50, 52)

(iii) 'Liberum arbitrium, naturaliter attributum a creatori animae rationali, illa media vis est, quae vel intendi ad fidem, vel inclinari ad infidelitatem potest.' (dSL 58)

The least these passages prove is that voluntary action, for Augustine, involves the ability to choose to do x or not to do x- that is, the ability to do otherwise. (That they do not, in fact, establish very much more than this, will be argued in Chs.8-9.)

Augustine, then, requires the ability to do otherwise as a condition of uncompelled action. Given, further, that for him

voluntary action must be uncompelled, and that he thinks that no deed can be responsible, worthy of praise or blame, unless it is voluntary (points I made in Ch.6, §2-4), this conclusion reflects interestingly on his much-discussed doctrine that the original human freedom was *an ability not to sin*, whereas the human freedom of the Resurrection will be *an inability to sin*: 'Primum liberum arbitrium posse non peccare, novissimum non posse peccare' (dCD 22.30).

Does Augustine's 'non posse peccare' describe any state of real freedom? If it does, then one wants to ask, with J.L.Mackie, why God did not give Adam *this* kind of freedom rather than the dangerous 'posse non peccare' variety. But it would seem that this eschatological freedom, as Augustine describes it, is not in fact much of a freedom. It is not clear that someone who is *unable* to sin is capable of *voluntarily refraining* from sinning. Hence it is also unclear that such an agent is responsible, or praiseworthy, for so refraining. After all, even a fallen human normally displays inability to sin in some respect or other. In my present state I cannot commit the sin of drunkenness, because I cannot afford it; nor the sin of fornication (as opposed to adultery), because I am married. Am I to be *praised* for my economically enforced sobriety, or my logically enforced non-fornication? Hardly; but then why should *any* kind of enforced sinlessness be praiseworthy? (Similarly, a theological point: if God's necessary sinlessness means that God *cannot* do wrong, why praise God for never doing wrong?)

5. The Cause(s) of 'Voluntas' Again

I began with the question: What, for Augustine, counts as an efficient cause of 'voluntas'? I noted the ambiguity of 'voluntas' between 'voluntary action' and 'volition', and a logical problem about seeing volitions as causes of actions. I considered two sub-conditions which Augustine gives upon the origination of voluntary action: that it should originate *with the agent*, and that it should originate *in such a way as to make it possible that it should have originated otherwise*. All this might seem insufficiently informative. How much does it tell us about how voluntary actions *do* in fact originate? We might wish to come back to the opening question: What does Augustine think the efficient cause of voluntary action is (if there is one)?

The question, of course, is a blunt instrument, and to do justice to the complexity of Augustine's thought on this issue, we need a rather more nuanced account of the matter than has yet been given. We must keep several questions firmly separate (as Augustine does not). We must distinguish (1) the question of the efficient cause of *volition*, from (2) the question of the efficient cause of *voluntary action*. We must also distinguish (1) and (2) from (3, 4...) questions about *other kinds* of cause.

On (1) and (2): as we have seen, many Augustinian texts strongly suggest that *volition* is the efficient cause of voluntary action. This idea has already been exposed to doubt. If, as I have suggested, a volition is not individuable as

anything more than an (unsuccessful) attempt to do a voluntary action, then the efficient cause of voluntary action will not be volition. On the contrary, volition and voluntary action will have the *same* kind of efficient cause (if they have any), which will be identical with neither of them.

But: what *is* this kind of cause- if there is one? Augustine's remarks on this question are ambiguous. In some places, as we have seen, his view is apparently that the 'voluntas' is altogether uncaused. In others, he seems to take a second view: that the 'voluntas' can be caused, but only by itself. In others still, he seems to argue for a third view: that there *can*, in fact, be causes of the 'voluntas' other than the 'voluntas' itself, and that an interesting theory of the causation of voluntary action can be elaborated. This third view is, I will argue in Chs.8-9, Augustine's most convincingly argued view about normal voluntary actions; although there is one special kind of voluntary actions to which the first and second views might seem to apply better.

For the first view, we might cite this evidence:

'Improba voluntas malorum omnium causa est... Quae tandem esse poterit ante voluntatem causa voluntatis? Aut enim et ipsa voluntas est; et a radice ista voluntatis non receditur: aut non est voluntas; et peccatum nullum habet. Aut igitur ipsa voluntas est prima causa peccandi, aut nullum peccatum est prima causa peccandi.' (dLA 3.48-9)

Augustine seems here to be arguing thus:

1. All possible causes of voluntary actions are themselves either voluntary actions or not voluntary actions.
2. 'X causes Y' = 'X gives necessary and sufficient grounds for Y's occurrence' = 'X compels Y'.
3. If a supposed voluntary action A had a cause C which was not itself a voluntary action, then A could not be a voluntary action, because A would have been compelled by C. Therefore no possible cause of a voluntary action could itself be other than a voluntary action.
4. And if a supposed voluntary action A had a cause C which was itself a voluntary action, then A could not be a voluntary action either, because A would have been compelled by C. Therefore no possible cause of a voluntary action could itself be a voluntary action.
5. But there are no other alternatives. Therefore voluntary actions have no causes.

From elsewhere, however, it seems that Augustine cannot be arguing that 'voluntas' has no cause. Augustine's point at dLA 3.48-9 is (as I argued in Ch.5, *ad* dLA 3.47 ff.) not about *all* voluntary action. He is *not* saying here that 'voluntas' (in general) has no prior cause. He is saying that '*mala* voluntas' (in particular) has no prior cause- indeed, that it is *only* '*mala* voluntas' which has no cause: other kinds of voluntary action do indeed have causes.

For the same reason, Augustine cannot be committed either to the view that nothing but one 'voluntas' can cause any other 'voluntas': even though he himself might be read as giving a lot of space, particularly in the dLA, to developing the view that 'voluntates' *typically* are the causes of 'voluntates'. The first pointer for this theory would be seen in the last passage quoted, if that quote was read in a different way:

'Quae tandem esse poterit ante voluntatem causa voluntatis? Aut enim et ipsa voluntas est; et a radice ista voluntatis non receditur...' (dLA 3.49)

This form of the 'reflexive theory of the "voluntas"' deserves a little further consideration—before its final rejection.

6. A Reflexive 'Voluntas'?

We have seen that, for Augustine, the cognitive abilities of the human mind are *reflexive*: to know (or perceive) anything is also to know that I know it (or perceive it). A foundational epistemological certainty is had by simple consciousness of one's self-consciousness. That I am a self-conscious living existent can, it is claimed, be established by simple introspection (dLA 1.16). An interesting parallel to this in Augustine's philosophy of action has appeared in the dLA. There, and sometimes elsewhere (e.g. dTrin 14.10), the nature of the human psyche in its

cognitive role is mirrored by its nature in its conative role. The structure of volition too is reflexive. This claim can be taken in different ways. I suggested in Ch.5 that it need imply no more than that 'bona voluntas' is *necessary* for the good life, because without it we cannot put anything else to good use. But it might also imply that 'bona voluntas' is *sufficient* for the good life: as I put it in Ch.5, "'bona voluntas" is had by simply choosing to have it. To *want* to have "bona voluntas" is *eo ipso* to have "bona voluntas". Clearly it is the second, stronger form of the reflexivity idea that we are interested in if we are considering the idea that the 'voluntas' is its own cause.

For (on the second reading) this theory makes the 'voluntas' its own cause, in the sense that the explanation of my having 'bona voluntas' of the first-order variety is that I made a (second-order) choice to have it. Presumably, if we are to look for an explanation of this second-order choice, we will be deferred to the *third* order; and so on *ad infinitum*. 'Voluntas', on this picture, is something absolutely independent of everything outside itself; something free standing, self motivating, and even, in something a little like Aristotle's sense, *self moving*.

The attractions of this theory to certain kinds of exegete of Augustine are obvious. If it is held that Augustine aims to argue away the threat of causal determinism, the reflexive theory of 'voluntas' seems to be a way of achieving that end. For, according to this theory, nothing causes the will except the will itself.

Indeed, it seems to be quite commonly thought that Augustine's aim is to present us with a *voluntarist* (as opposed to *rationalist*) theory of the voluntary, *i.e.* a theory based on a 'voluntas' which 'being anterior to reason, has at the most fundamental level no reason for its biddings' (Macintyre 1985, p.156). If the reflexive theory of the voluntary is Augustine's most developed thinking on the subject, then it seems possible to argue that, taking him all in all, Augustine is indeed a voluntarist in the sense required.

Of course, it would then have to be conceded that the explanation of the activity of the 'voluntas' offered by this theory is essentially regressive and uninformative. First-order choices are referred to second-order choices; second- to third-order choices; third- to fourth-; and so on. If the 'voluntas' is to be caused by *nothing* but the 'voluntas', then we cannot use the escape route from this regress suggested in Ch.5, and say that at some order there can be a 'voluntas' which is not caused by another 'voluntas'. Then the activity of the 'voluntas' will remain unexplained, and apparently inexplicable. As already suggested, the rejoinder might be that this inexplicability of the 'voluntas' is exactly what we ought to expect in a voluntarist world. If it is true (i) that the deliberations of the 'voluntas' are anterior to reason, and (ii) that the 'freedom' of the 'voluntas' means its being radically unconditioned by causes, then clearly (it might be said) a 'libera voluntas' is *bound* to be inexplicable.

To put this in a less favourable light: the reflexive theory seems to lead either to a regress or else to the admission that there is a first term of the regress which is a will which is not caused at all— thus establishing the same as the first view, the view that the 'voluntas' as such is uncaused. Some might think that this is how it ought to be. Others, in the light of much else that Augustine says, will view this prospect with suspicion. In particular, as we have already seen in Ch.5, in exegesis of dLA 3.47 ff., it is not at all clear that the exegetical basis for attributing the strong version of the reflexivity thesis to Augustine is sound: it cannot be if Augustine there means to talk only about '*mala* voluntas' and not about 'voluntas' in general. As I have argued that Augustine does indeed mean to talk only about '*mala* voluntas' at dLA 3.47 ff., my reasons for rejecting the strong version of the reflexivity idea should be plain.

One last point in the strong version's favour, however, may be mentioned before leaving the topic. Such a recursive understanding of 'voluntas' might seem to connect with, perhaps even to *support*, the circular idea of responsibility pointed to in the last chapter. The suggestion there was that fallen humans are culpable for the sins which they cannot help committing, because they are in a state of helplessness, which itself is the penalty of committing those same sins. The suggestion here is that the only possible cause of any *n*th-order 'voluntas' is an (*n*+1)th-order 'voluntas'. So the *moral quality* of any one 'voluntas' of any person is fixed, if by anything, only by another 'voluntas', which will (*ipso facto*) share that moral

quality. Ascriptions of praise and blame apply, therefore, equally to all the individual 'voluntates' of a given person. Thus our account of responsibility *has* to be circular; for our account of motivation is circular. The moral state in which we find ourselves is not something to be *explained*; it is simply a *given*, just as (as Ricoeur remarked, Ricoeur 1967) one point of the Adam and Eve story is that the prevenient existence of evil is a given even in the Garden of Eden.

However, if there is a connection to be made here, Augustine never explicitly made it. Although it is never formally repealed, the idea of the reflexivity of 'voluntas' is only developed in the early works. The idea of 'reflexive responsibility', on the other hand, is most commonly developed in the later works. The only work in which *both* ideas may be found is, as we have seen, the dLA; but no tie up is attempted by Augustine, which of course suggests that this was not what he had in mind.

Moreover, there are other, crucial developments in his doctrine which stand in the way of a *general* ascription to Augustine of this kind of reflexive theory of the voluntary. A broader look at Augustine's philosophy of action shows that it is, overall, very far from the truth to say that he believes that *typical* voluntary action is either uncaused or rationally undetermined. On the contrary: Augustine has a very definite and prominent theory of practical reasoning, according to which rational determination is in fact *characteristic* of voluntary action, and moreover is itself the *cause* of voluntary action. To

that theory, in the next chapter, I will now turn.

1. Introduction

2. The nature of the world and the nature of the mind

3. The nature of the world and the nature of the mind

4. Conclusion

5. The nature of the world and the nature of the mind

The nature of the world and the nature of the mind

The nature of the world and the nature of the mind

The nature of the world and the nature of the mind

The nature of the world and the nature of the mind

The nature of the world and the nature of the mind

The nature of the world and the nature of the mind

1. Introduction

The nature of the world and the nature of the mind

The nature of the world and the nature of the mind

The nature of the world and the nature of the mind

The nature of the world and the nature of the mind

The nature of the world and the nature of the mind

Chapter 8

The Good Will and the Good Life

1. Introduction
2. Practical Reason and Practical Wisdom in Augustine
3. Aristotle and Augustine on the Directedness of Action
4. 'Felicitas'
5. Good Will and the Order of the World

'Nulla est homini causa philosophandi nisi ut beatus sit.'

(Augustine, *City of God* 19.1)

'Quia fecisti nos ad te, et inquietum est cor nostrum donec requiescat in te.'

(Augustine, *Confessions* 1.1)

1. Introduction

Augustine has much to offer us in the way of demarcation of the conditions of voluntary action, as has become clear in Chs.6 and 7. He develops a number of (not always consistent) lines of thought on the subject. But among his principal themes are the very Aristotelian ones that absence of compulsion and of

ignorance are negative conditions for voluntary action; and that among the positive conditions for voluntary action are (i) that the action should originate with me and (ii) that I should be able to do otherwise.

For my analysis of these conditions I took as a starting point the first two phrases of Augustine's description of voluntary action, at dDA 14, as an 'animi motus, cogente nullo, ad aliquid vel non amittendum vel adipiscendum'. In this chapter I turn to the last phrase of this definition. The phrase 'ad aliquid vel non amittendum vel adipiscendum' points us to a question which is also suggested by the comparison with Aristotle: what does Augustine have to say about the *rationality* of voluntary action?

Of Aristotle's theory I noted that his three conditions for voluntary action can most succinctly be expressed in their negative forms, as the requirements that voluntary action should not be (i) compelled, (ii) done in ignorance, or (iii) irrational. But I also noted that Aristotle says a great deal to fill out these negative conditions with a positive content. In particular, it is a positive consequence of Aristotle's negative condition about irrationality that fully voluntary action must logically follow from the combination of a premiss of the good with a premiss of the possible.

We have seen that there are correlates, in Augustine's theory of the voluntary, to Aristotle's conditions, positive and

negative, about compulsion and ignorance. The question I ask in this chapter is: Does Augustine have anything corresponding (negatively) to Aristotle's 'no irrationality' requirement, or (positively) to his account of the role of practical reason in voluntary action?

2. Practical Reason and Practical Wisdom in Augustine

My answer is a very guarded 'Yes', with two *caveats* in particular, (i) and (ii) as noted below. Contrary to what many of his interpreters think, Augustine is no irrationalist: he does not believe that it is normal that nothing should motivate human choices except the sheer fiat of an inscrutable 'black box' called, say, 'the will'. Genuine voluntary action, for him just as for Aristotle, is action on a good reason:

'Pax animae irrationalis ordinata requies appetitionum, pax animae rationalis ordinata cognitionis actionisque consensio.'
(dCD 19.13)

'Sed quia homini rationalis anima inest, totum hoc, quod habet commune cum bestiis, subdit paci animae rationalis, ut mente aliquid contempletur et secundum hoc aliquid agat, ut sit ei ordinata cognitionis actionisque consensio.'
(dCD 19.14)

Note the contrast between 'anima rationalis' and 'anima irrationalis': a contrast which might have been taken straight

from NE 1102a29-30 (though more probably it came from a Platonist writer). If Augustine really believed, like David Hume, that the choices of a human soul were incapable of receiving the direction of reason- and that in fact the relation of control ran the other way- we might be surprised to find Augustine saying, as he so often does, that the human soul is, by nature, 'rational' (whatever precise meaning that word may have).

For Augustine as much as for Aristotle, genuine voluntary action involves the *congruence* ('consensio') of one's action with one's knowledge and one's desire; or, to put it another way, it entails that one's action should, in some appropriate sense, *follow from* the combination of one's knowledge and one's desire. What is this if not an account, or at any rate the outline of an account, of practical reasoning?

However, (i) there is (as we shall see) one large problem with describing Augustine's account of practical wisdom as a *condition* of voluntary action, since Augustine emphasises that wilful wrongdoing, action directly against practical wisdom, is not necessarily involuntary. The complexities raised by this problem will be considered in Ch.9.

Also (ii), unlike Aristotle in NE III and VI, Augustine nowhere fills in the outline with a detailed, explicit account of practical reason. This is not to say that he does not fill in that outline: it is to say that he fills it in with something else. Aristotle offers us an account of the mechanics of

practical reasoning which, I have suggested, is virtually a formal logical account, and can certainly be presented as such with some tidying up. Augustine, on the other hand, concentrates on the other side of the picture. The mechanics and the logic of practical *reason* he leaves almost entirely implicit and unexplored, though this is no excuse for saying that he has no such concept as practical reason. But on the more nebulous matter of practical *wisdom*, Augustine has plenty to say.

Practical wisdom, for Augustine, means understanding what the good for humanity is so as to live it out. Discovering the good for humanity in one's own life is not just a result of deliberating well (which is why this kind of practical understanding is more appropriately called wisdom than reason). A kind of *revelation*, a kind of special *knowledge*, is needed. (V. dTrin 13.4, 'Non quod aliquis [beatitudinem] *no*lit, sed quod non omnes eam *no*rint'; cp. dFRV 1).

On the other hand, Augustine often speaks of this special knowledge as if it were completely deducible on the grounds of reason alone (as in the 'proof' of God's existence at dLA 2.7-39). This is an example of a standard tension in Christian apologetics, between natural and revealed theology; Platonic $\gamma\upsilon\omega\sigma\iota\varsigma$ displays the same ambiguity of status between the natural and the revealed. It is in accord with that ambiguity that Augustine's quest for the good life begins with the frankly phenomenological just about as often as it begins with the transcendently metaphysical. I consider both approaches.

2a. The Good Life: Phenomenological Arguments

On the phenomenological side, the maxim of Augustine's practical wisdom is the Delphic γνῶθι σεαυτὸν. The method of Augustine's work often suggests that practical wisdom is no more than self awareness: becoming aware of *what I really want*, the better to pursue it. It means knowing what is the natural objective of all action and desire. We can arrive at practical wisdom by discovering what this objective is. And one way to discover that objective, since it is natural, is by simply examining what humans (for example, me) actually do pursue, what they intend to gain by this pursuit, and what the difference is, if any, between practice and underlying intention.

This examination is performed at the beginning of several of Augustine's works, e.g. the dBV, the cAcad, and the dLA. What such an examination finds is, unsurprisingly, a great variety of different objectives, and a great variety of degrees of success or failure in humans' attempts to achieve what they are seeking to achieve. 'Quomodo enim voluntate quisque miseram vitam patitur, cum omnino nemo velit misere vivere?' (dLA 1.30, cp. dTrin 14.4): how do we explain this gap between intention and performance? Augustine's answer is that it demonstrates that not just any route will lead one to the happy life:

'Ad hoc pergendo, quod aut non est, aut, si est, non facit beatos, ad beatam vitam nullus pervenire potest.' (dLA 3.59)

This is where the need for practical wisdom appears. There is a gap between wanting the good and pursuing the good:

'Nam illi omnes quos commemorasti diversa sectantes, bonum appetunt et malum fugiunt; sed propterea diversa sectantur, quod aliud alii videtur bonum. Quisquis ergo appetit quod appetendum non erat, tametsi id non appeteret nisi ei videretur bonum, errat tamen.' (dLA 2.26)

It is right to pursue the good life, yet there are wrong, unwise, ways to pursue it:

'Inquantum igitur omnes homines appetunt vitam beatam, non errant. Inquantum autem quisque non eam tenet vitae viam quae ducit ad beatitudinem, cum se fateatur et profiteatur nolle nisi ad beatitudinem pervenire, intantum errat... *Et quanto magis in via vitae quis errat, tanto minus sapit.*' (dLA 2.26)

Phenomenological examination of our natural desires shows that '[Nemo] beatus [est] qui quod vult non habet' (dBV 10). Unhappiness is caused by seeking without finding (cAcad 1.6-8), or by finding and then losing again. One should therefore seek the objectives of greatest permanence for the maximal satisfaction of desire. Wrong objectives characteristically do not admit of that 'stable and permanent possession' which Boethius talks about as characteristic of eternal happiness (Boethius, *The Consolation of Philosophy*, Bk.V, Prose 6). So in the dLA Augustine argues, rather implausibly, that all and only

bad objectives are transient:

'Cupere namque sine metu vivere, non tantum bonorum, sed etiam malorum omnium est: verum hoc interest, quod id boni appetunt avertendo amorem ab iis rebus, quae sine amittendi periculo nequeunt haberi; mali autem ut his fruendis cum securitate incubent, removeere impedimenta conantur, et propterea facinorosam sceleratamque vitam, quae mors melius vocatur, gerunt.' (dLA 1.10)

If Augustine can make out this step of the argument (more on which below), that the right objectives of action are the most permanent ones, that will bring us to the metaphysical form of his argument about the objectives of practical wisdom.

2b. The Good Life: Metaphysical Arguments

On the metaphysical side, practical wisdom is the possession of the highest good in virtue of understanding a certain kind of *truth*, the truth about what is *really* the 'summum bonum', independently of and antecedently to the various views of it which may be held. So dLA 2.36:

'Haec enim veritas ostendit omnia bona, quae vera sunt, quae sibi pro suo captu intelligentes homines vel singula vel plura eligunt, quibus fruuntur.'

Action- which is to say pursuit of some perceived good- needs to follow from genuine knowledge of what is worth pursuit. Right cognition, cognition of the truth, is a condition of successful voluntary action. True to his Platonist roots, Augustine stresses the role of knowledge in practical wisdom. For him as for Socrates 'virtue is knowledge'- at least in the sense that knowledge is necessary for all virtue, even if it is not also sufficient.

Moreover, what we want for practical wisdom is not just the truth given by self examination; it is truth in a rather loftier sense, a truth identical with wisdom and common to all humans:

'Num aliam putas esse sapientiam nisi veritatem, in qua cernitur et tenetur summum bonum?' (dLA 2.26)

'Si summum bonum omnibus unum est, oportet etiam veritatem in qua cernitur et tenetur, id est sapientiam, omnibus unam esse communem.' (dLA 2.27)

(This last passage, incidentally, suggests a response to a likely objection to Augustine's doctrine of practical wisdom. It may be said that Augustine's argument is fatally flawed because he illegitimately infers from 'All humans seek some good thing in all their actions' to 'There is some (one) good thing which all humans seek in all their actions'. If Augustine argued like this, certainly he would be guilty of a logical fallacy. But in fact, he argues the other way round. He starts from the claim that

there is one good at which all action aims. So the above passage begins from this assumption: 'Si summum bonum omnibus unum est...'. He does not deduce this claim from the claim that all human action aims at some good or other. On the contrary, that latter claim is deduced from the first claim, by a logically unimpeachable transition.)

So lofty is Augustine's metaphysical conception of practical truth that, for him, practical truth turns out to be exactly the same in content as his theoretical truth. It is natural to all men to seek the (apparent) good and flee the (apparent) evil (cp. dLA 1.30, 'nam hoc volunt etiam mali'). But what more is necessary to the good life is exactly the same knowledge of transcendent and objective reality as forms the end of theoretical enquiry. In this sense Augustine's wisdom is as unitary as the objective truth which it contemplates: in fact, there is no neat division between practical and theoretical wisdom in his thought. Just as, for Augustine, talk of 'voluntas' and 'ratio' relates to the same human person in the different roles of agent and percipient, so also, for him, *practical* reason is only theoretical reason put to work, and *theoretical* reason is only practical reason in contemplative mode. The first principles of practical and theoretical reason are, in his view, one and the same.

'Sic fortis acies mentis et vegeta cum multa vera et incommutabilia certa ratione conspexerit, dirigit se in ipsam veritatem, qua cuncta monstrantur.' (dLA 2.36)

This conjunction of metaphysics and phenomenology is also a conjunction of moral and factual claims. Augustine holds *both* that our desires naturally *have* a certain form, and that our desires *ought* to have that form. The vision of the good for humanity is, on the one hand, a kind of θεωρία of what is objectively the case. But, on the other hand, it is also action guiding, indeed it is preeminently so. Augustine's endorsement of the idea that voluntary action is necessarily, in some sense, *reasonable* or *rational* behaviour is an endorsement of the claim, also made by Aristotle, that voluntary action has an intrinsic *directedness*, namely towards objectives which *necessarily* are always either actual or at least perceived *goods*.

3. Aristotle and Augustine on the Directedness of Action

This notion of directedness is a crucial one, and needs further elucidation. Before turning to Augustine's statements of this notion, I will briefly review what Aristotle meant by it.

It would not make sense to Aristotle for someone to say that the reason why they wanted, say, to drink a can of paint was 'because it seemed like a bad thing to do'. Purported explanations of an action, which do not refer to some real (or supposed) *good* which is obtained (or thought to be obtained) by that action, are simply not explanations at all.

Thus, for Aristotle, it makes sense to explain 'John wants to

drink a can of paint' by '...because he has made a bet about it'; or '...because he is being forced to at gunpoint by a maniac'; or '...because he wrongly imagines that the can of paint is a can of beer'; or even merely '...because he thinks it would be a neat thing to do'. Even this last could count as a rational explanation of an action, in the sense of 'rational' which I have in mind. But it does *not* make sense to explain 'John wants to drink a can of paint' by '...because he hates the taste of paint', or '...because he knows that paint is poisonous', or '...because he thinks it would *not* be a neat thing to do'. As they stand, these 'explanations' simply aren't explanations of John's behaviour at all. Consider, for one thing, how much more naturally they would read if, in each one, 'although' were substituted for 'because'.

Of course, we can easily adorn all three non-explanations so that they *do* read like pukka explanations. E.g., we can add to them, respectively, '...and wants to inure himself to horrible tastes', '...and is trying to commit suicide', and '...and is aiming to disgust his girlfriend into leaving him'. But what is the point, the attractiveness, of adding these adornments, if not that they restore intelligibility to our characterisation of John's behaviour *precisely by indicating a conceivable good which it aims at*? Which is simply another way of making the same point: that Aristotle's teaching is emphatically that 'Quidquid appetitur, appetitur sub specie bonitatis' (as Aquinas puts it).

Now, as we have seen, Augustinian voluntary action is not

primarily rational in the tight sense that Augustine too has an explicit, closely argued account of practical reasoning to offer us. Rather, Augustinian voluntary action is principally to be called 'rational' in the looser sense that he too holds that *no behaviour is (fully) explicable as voluntary action unless we can give some account of the good at which it aims or is supposed by the agent to aim*. For Augustine too, to explain a piece of behaviour in terms of *reasons* (as opposed to other types of causes) is to mention a *good* which the behaviour is supposed to be conceived as aiming at by the agent. That this is a belief of Augustine's, and such a fundamental one that much of the time it remains unexamined, is already evidenced by one casual aside I have quoted, from dLA 2.26: '*Quisquis ergo appetit quod appetendum non erat, tametsi id non appeteret nisi ei videretur bonum...*'. Here Augustine takes it absolutely as read that no one would pursue anything unless they held it, for one reason or another, to be a good objective to pursue.

If there is evidence of Augustine's endorsement of the thesis of the directedness of voluntary action, then a most important tie up can be made in Augustine's moral theory: between *practical rationality* and *practical virtue*, between what it is *rational* or *reasonable* to do, and what it is *good* to do. But the evidence is no accumulation of slips of the pen; it is Augustine's consistent doctrine. Further evidence of that doctrine can be found by examining Augustine's expression of the thesis of the directedness of voluntary action in his theory of 'felicitas' or 'beatitudo'.

4. 'Felicitas'

The central importance to Augustine's thought of the question 'What is the good/ best life for humanity?' is obvious from one end of his philosophical career to the other:

'De beata vita [,Theodore,] quaesivimus inter nos, nihilque aliud video quod magis Dei donum vocandum sit.' (dBV 1.5)

'Nam cum beati esse cupiamus, sive id fieri non potest nisi inventa, sive non nisi diligenter quaesita, veritate; postpositis ceteris omnibus rebus, nobis (si beati esse volumus) perquirenda est [veritas].' (cAcad 1.25)

'Socrates animum [intendebat], quod esset beatae vitae necessarium, propter quam unam omnium philosophorum invigilasse ac laborasse videtur industria...' (dCD 8.3)

'Quid quod et Felicitas dea est?... Ipsa ergo sola coleretur. Ubi enim ipsa esset, quid boni non esset?' (dCD 4.18)

How does Augustine answer this question of the nature of the good/ best life? Like Aristotle (NE 1097b22 ff.) and like Cicero in the *Hortensius*, he generally begins with what is often thought to be a truism: that the good for humanity is happiness.

'Beati certe omnes esse volumus.' (dBV 10; cp. cAcad 1.5, Sermon 150.4, dTrin 13.4 and Cicero, *Hortensius*, Fr.36 Müller)

'Cum ergo beati esse omnes homines velint...' (dTrin 13.4)

'Quis enim optat aliquid propter aliud quam ut felix fiat?'
(dCD 4.23)

'Nemo est qui gaudere nolit.' (dCD 19.12)

'Utique peccando nec pietatem nec felicitatem tenuimus,
voluntatem vero felicitatis nec perdita felicitate perdidimus.'
(dCD 22.30)

But ποθεῖται δ' ἐναργέστερον τί ἐστίν ἔτι λεχθῆναι (NE 1097b23): what does it actually *mean* to say that 'laetitia', 'felicitas', 'beatitudo' is the good for humanity? Augustine has at least seven ways of approaching this question, which I will now run through. These seven approaches may or may not come down to the same thing, as they are intended to; and they may or may not depend circularly on each other, as they are not intended to.

4a. Peace

One standard image for the human good in the *City of God* is that of the final rest of the saints in 'pax' or 'quies':

'[Possumus] dicere fines bonorum nostrorum esse pacem.' (dCD 19.11)

'Sicut enim nemo est qui gaudere nolit, ita nemo est qui pacem habere nolit.' (dCD 19.12)

Is this peace the same to which Augustine had referred in the schema of moral progress noticed (Ch.6, §4) in his exegetical works on Romans and Galatians ('ante legem', 'sub lege', 'sub gratia', 'in pace')? There is good evidence in the dCD that it is. The point about the 'pax' referred to in the exegetical works is that such 'pax' only comes after the end of the struggle between good and evil impulses- which itself is characteristic of unhappiness. The 'pax' of the dCD is precisely this kind of absence of internal conflict. For a rather unfairly unsympathetic criticism of Augustine's doctrine of 'pax' as the final good, v. Kirwan (1989), p.222 ff.: 'It is striking that apart from the reference to worship this description at the end of the *City of God* is wholly negative'. But in fact the absence of strife in heaven is an absence which, as we have seen, Augustine identifies in the dCD with something very positive, viz. the possibility of successful practical reasoning:

'Pax animae irationalis ordinata requies appetitionum, pax animae rationalis ordinata cognitionis actionisque consensio.' (dCD 19.13)

One notable point about defining the human good as 'pax' in this sense is that, given Augustine's orthodox Christian belief in human non-perfectibility in this life, it makes that good essentially other-worldly. The point is clear already in the

EPRoms and EEGal:

'Non finientur haec [desideria carnalia] nisi resurrectione corporis immutationem illam, quae nobis promittitur, meruerimus, ubi perfecta pax erit.' (EPRoms 18)

'Postea vero ex omni parte exstinguitur [poenalis consuetudo]. Quoniam Spiritus Iesu... vivificabit mortalia corpora nostra.' (EEGal 46)

Our lot in the present life is wretchedness and punishment, as is increasingly strongly emphasised as Augustine's thought develops:

'Omnes homines, quamdiu mortales sunt, etiam miseri sint necesse est.' (dCD 9.15)

'Hanc vitam de peccato illo nimis nefario, quod in paradiso perpetratum est, factam nobis esse poenalem, totumque quod nobiscum agitur per testamentum novum, non pertinere nisi ad novi saeculi hereditatem novam.' (dCD 21.15)

God is only to be known in the hereafter. The 'pax' or 'quies' for which we long is only to be found in the perfected knowledge of God, which Augustine also calls

4b. The contemplation of God

'Quia fecisti nos ad te, et inquietum est cor nostrum donec requiescat in te.' (Conf 1.1)

'Ibi videbimus, et gaudebit cor nostrum [Isaiah 66.14]. Nec expressit [Esias] quid videbimus: sed quid nisi Deum? Ut impleatur in nobis promissum evangelicum: Beati mundicordes, quoniam ipsi Deum videbunt [Mt.5.8]... Hic creditis, ibi videbitis.' (dCD 20.21)

Since this contemplation of God is happiness, the lack of it is *unhappiness*:

'Illa namque visio Dei tantae pulchritudinis visio est et tanto amore dignissima, ut sine hac quibuslibet aliis bonis praeditum atque abundantem non dubitet Plotinus [*Enneads* 1.6.7] infelicissimum dicere.' (dCD 10.16)

4c. Function

Very different in tone, and implicitly much more this-worldly, are Augustine's arguments about *function*, in the performing of which one's happiness is often, especially in the early works, said by him to consist. Like Aristotle, he takes 'human function' to mean 'activity/ state which sets humans "above" the animals':

'Quid censes, inquam, esse aliud beate vivere, nisi secundum id quod in homine optimum est vivere?... Quis, inquam, dubitaverit nihil aliud esse hominis optimum, quam eam partem animi cui dominantī obtemperare convenit cetera quaeque in homine sunt?' (cAcad 1.5; cp. Aristotle, NE 1113a5-9)

'Hoc quidquid est quo pecoribus homo praeponitur, sive mens, sive spiritus, sive utrumque rectius appellatur... Si dominetur atque imperet ceteris quibuscumque homo constat, tunc esse hominem ordinatissimum.' (dLA 1.18)

It follows that *not* performing the human function is often taken to be definitive of unhappiness; as when one performs some other, lower animal's function:

'Pulchre namque incedit quadrupedans equus: at si hoc homo pedibus manibusque imitetur, quis eum vel palarum cibo dignum putet? Recte igitur plerumque improbamus imitantem, cum eum quem imitatur probemus.' (dDA 20)

-Or as when one's own body is disobedient to one's own 'voluntas'- a phenomenon which seems to have fixated Augustine (dCD 14.23). In heaven the body will, he says, no longer disobey the 'voluntas'. As evidence of this he adduces the fact that even now some people can control their bodies in remarkable ways, and goes on:

'Cum itaque corpus etiam nunc quibusdam, licet in carne

corruptibili hanc aerumnosam ducentibus vitam, ita in pleribus motionibus et affectionibus extra usitatum naturae modum mirabiliter seruiat; quid causae est, ut non credamus ante inoboedientiae peccatum corruptionisque supplicium ad propagandam prolem sine ulla libidine servire voluntati humanae humana membra potuisse? Donatus est itaque homo sibi, quia deseruit Deum placendo sibi, et non oboedens Deo non potuit oboedire nec sibi. Hinc evidentior miseria, qua homo non vivit ut vult.’ (dCD 14.24)

The obvious response to ‘The happy life for humans is the life of living out the human function’ is ‘Yes, but what *is* the human function?’. The above passages give some indications about what Augustine thinks the human function is. More evidence as to his beliefs on this may be gathered by examining another way in which he describes the good life, as

4d. Desire of the right things

The argument that the good life is a matter of having the right desires, and (*if* one has the right desires) of fulfilling them, is very clearly presented in the dBV. It is agreed there that happiness has to do with the satisfaction of desires:

‘Videturne vobis, inquam, beatus esse qui quod vult non habet? Negaverunt.’ (dBV 10)

But not just of *any* desires:

'Quid? omnis qui quod vult habet beatus est? Tum mater: Si bona, inquit, velit et habeat, beatus est; si autem mala velit, quamvis habeat, miser est.' (dBV 10)

Compare the teaching of the *de Trinitate*:

'Velle enim quod non deceat, idipsum miserrimum est; nec tam miserum est non adipisci quod velis, quam adipisci velle quod non oporteat.' (Cicero, *Hortensius*, *apud* dTrin 13.5)

What, then, are the right things to desire? Augustine gives us these examples of *wrong* objectives of desire, desires which could not be given a central part in the constitution of the good life for humans:

'Sunt alia quaedam, quae iam cadere in feras non videntur, nec tamen in homine ipso summa sunt, ut iocari et ridere; quod humanum quidem, sed infimum hominis iudicat, quisquis de natura humani rectissime iudicat. Deinde amor laudis et gloriae, et affectatio dominandi...' (dLA 1.18)

'Unde illa cuiusdam mimi facetissima praedicatur urbanitas, qui cum se promisisset in theatro quid in animo haberent et quid vellent omnes, aliis ludis esse dicturum, atque ad diem constitutum ingenti expectatione maior multitudo conflueret, suspensis et silentibus omnibus dixisse perhibetur: Vili vultis emere, et caro vendere.' (dTrin 13.3)

'Corporea diligendo peccamus, quia spiritualia diligere et iustitia iubemur, et natura possumus.' (dDA 20)

Augustine has three positive answers to the question of what right desire is. These answers seem different in content; different enough, at least, to supply two further ways of answering the earlier question, 'What is the good life?'. First, in line with (54b), that the good for humanity is the contemplation of God, he argues that the right supreme object of desire is God himself. Second, it seems in many places that the right thing to desire is not the happy life, but

4e. Righteousness

'Nam illi qui beati sunt- quos bonos esse oportet- non propterea sunt beati, quia beate vivere voluerunt (nam hoc volunt etiam mali), sed quia recte, quod mali nolunt.' (dLA 1.30)

'Nisi beatus non vivit ut vult, et nullus beatus nisi iustus.' (dCD 14.25)

'Recta vita ducenda est, qua perveniendum sit ad beatam.' (dCD 14.9)

It appears that this right kind of life is defined as that life which exhibits all the emotions in a right kind of way, and a misdirected life is one which exhibits them in the wrong way

(dCD 14.9): 'Quia rectus est amor eorum, istas omnes affectiones habent'.

4f. Secure Goods

The third answer found in the dLA is: we should desire those things *which we cannot lose against our will*. This is the difference between the right desires of good people and the wrong desires of bad people:

'Cupere namque sine metu vivere non tantum bonorum, sed etiam malorum omnium est. Verum hoc interest, quod id boni appetunt, avertendo amorem ab iis rebus quae sine amittendi periculo nequeunt haberi; mali autem ut his fruendis cum securitate incubent, remove impedimenta conantur, et propterea facinorosam sceleratamque vitam, quae mors melius vocatur, gerunt.' (dLA 1.11)

And what, in turn, is it that we cannot lose against our will? In the dLA Augustine answers that what we cannot lose against our will is

4g. Bona Voluntas

-The good will itself. This is the 'voluntas qua appetimus recte honesteque vivere, et ad summam sapientiam pervenire' (dLA

1.25). Its value is greater than 'anything in the way of riches, or honour, or the desires of the body' for the simple reason that one has good will by choosing to have it, and can only lose it by choosing not to have it. This makes the good will the most secure of all goods, a good completely independent of the ravages of fortune and time:

'Vides igitur iam in voluntate nostra esse constitutum, ut hoc vel fruamur vel caremus tanto et tam vero bono. Quid enim tam in voluntate quam ipsa voluntas sita est?' (dLA 1.26)

'Placet igitur beatum esse hominem dilectorem bonae voluntatis suae, et prae illa contemnentem quodcumque aliud bonum dicitur, cuius amissio potest accidere etiam cum voluntas tenendi manet.' (dLA 1.28)

The same kind of thought is found in the dCD: 'Beata quippe vita si non amatur, non habetur' (dCD 14.25).

Now there are at least two important problems with defining the human good in this manner, as the 'bona voluntas'. Firstly, the security of possession of the 'bona voluntas' is no guarantee of its goodness as an object of possession. After all, by the same arguments that Augustine gives for the security of possession of the 'bona voluntas', one could argue that the '*mala voluntas*', if there is such a thing, was a secure possession- and so a good thing? The argument from permanence of a possession to its goodness is implausible because, as Augustine himself points

out, as early as the dBV, that desire for or possession of something is not a good state of affairs unless that something is a *good* thing:

'Si bona [quisquam] velit et habeat, beatus est; si autem mala velit, quamvis habeat, miser est.' (dBV 10)

(Why, incidentally, shouldn't the pursuit and possession of wrong objectives make one happy? One standard answer of Augustine's is that the satisfaction to be had in the attainment of bad things is not a real satisfaction. In truth it makes one more miserable than happy to obtain such things: 'Quis namque ita sit mente caecus... ut eum qui nequiter vivit ac turpiter, et... implet omnes suas facinorosissimas et flagitiosissimas voluntates, ideo beatum dicat, quia vivit ut vult: cum profecto quamvis et sic miser esset, minus tamen esset, si nihil eorum quae perperam voluisset, habere potuisset?' (dTrin 14.8). Once again, the argument is of fundamentally phenomenological form: Augustine's point is that even the wicked know, in their heart of hearts, that what they pursue and obtain is not *really* of a kind to bring them happiness. Though of course if they failed to know this, that ignorance would be just another sign of their iniquity; they lose either way.)

Correspondingly, as Augustine remarks in the *Confessions* (6.16), their lack of permanence is, on its own, no argument against the supremacy of the 'carnal' goods. For the implication seems to be that the 'carnal' goods might be the supreme goods,

if they *did* last for ever; but Augustine hardly wants to concede that. (The tendency to appeal, irrelevantly, to the longevity of a good as evidence of its 'true' goodness is a chronic vice of Platonism, which Aristotle's tart medicine should really have cured once and for all: ἀλλὰ μὴν οὐδὲ τῷ ἀϊδίῳ εἶναι μᾶλλον ἀγαθὸν ἔσται, εἴπερ μηδὲ λευκότερον τὸ πολυχρόνιον τοῦ ἐφημέρου (NE 1096b3-5).)

Secondly: for a Christian bishop, there might seem a danger of some kind of unorthodoxy or idolatry in saying that the greatest good for humanity is 'bona voluntas', and that we should love this 'voluntas' and scorn *everything* (!) else. After all, where does *God* enter this story? In the section of the *Retractationes* dealing with the dLA (Retr 1.8), Augustine does not explicitly repudiate this claim. But he does point out that even in the dLA the 'potentiae animi', including 'bona voluntas', are only 'media bona' (even if 'tam vera bona'), and that he there stressed the need for divine grace to assist human 'bona voluntas':

'Voluntas ergo ipsa nisi Dei gratia liberetur a servitute, qua facta est "serva peccati", et ut vitia superet adiuvetur, recte pieque vivi a mortalibus non potest.' (dLA 1.11)

This suggests that possession of the 'bona voluntas' alone is not, in fact, sufficient for the possession of the good life. God's grace is also needed. But the supplying of that grace is clearly not 'in voluntate nostra', unless our possession of the

'bona voluntas' guarantees God's supplying of divine grace- a suggestion which Augustine and a host of followers have strenuously denied.

More usual answers to the question of the nature of the human good, in the later works especially, are the ones reviewed already, that it is (4a) peace or (4b) contemplation of God. Augustine's doctrine of the 'bona voluntas' was never meant to be a denial of these answers, even if it looks rather like one at times. As more careful examination of the dLA shows, the 'bona voluntas' is not so much the greatest good for humans as *the desire for* (or choice of) the greatest good for humans. It is therefore the best desire to have and the best choice to make, but this is only because it is desire/ choice of the best possible objective- which in Augustine's view is of course God. Once again, we see the need to adopt a relatively modest view of the role of 'voluntas' in Augustine's thought. The best interpretation of Augustine is not that the possession of 'bona voluntas' is sufficient for the good life, but only that it is necessary.

Formally speaking, Augustine's way of reaching the conclusion that the best kind of 'voluntas' is the desire for God is to revert to the 'function' style of argument. It is the human function to turn to what is good. What is good? What is eternal:

'Iubet igitur aeterna lex avertere amorem a temporalibus, et eum mundatum convertere ad aeterna.' (dLA 1.32)

What is good is also, characteristically, not private, it is equally available to all. This is a principle of metaphysics in the dLA:

'Satis enim est quod istas tanquam regulas, et quaedam lumina virtutum, et vera et incommutabilia, et sive singula sive omnia communiter adesse ad contemplandum eis qui haec valent sua quisque ratione ac mente conspicere, pariter mecum vides certissimumque concedis.' (dLA 2.29)

In the dCD it is a principle of political philosophy too:

'[Civitas libera erit civitas Dei], ubi sit non amor propriae ac privatae quodam modo voluntatis, sed communi eodemque inmutabili bono gaudens atque ex multis unum cor faciens.' (dCD 15.3)

'Quanto magis homo fertur quodam modo naturae suae legibus ad ineundam societatem pacemque cum hominibus... omnibus obtinendam.' (dCD 19.12)

Now what is common and eternal in this way, except for truth itself?

'Promiseram autem... me tibi demonstraturum esse aliquid quod sit mente nostra atque ratione sublimius. Ecce tibi est ipsa veritas; amplectere illam si potes, et frui illa... Quid enim petis amplius quam ut beatus sis? Et quid beatius eo qui fruitur

inconcussa et incommutabili et excellentissima veritate? ...An nos in amplexu veritatis beatos esse dubitamus?' (dLA 2.35)

In the very Plotinian *de Quantitate Animae*, the 'contemplatio veritatis' is made the seventh and last stage in an 'ascensio animae' from minimal vitality, *via* sensation, technical knowledge, the work of purgation, moral purity, and the hunger for What Is: 'septimus atque *ultimus* gradus', with nothing better to follow it, for there is nothing better than the 'perfructio summi et veri boni' (dQA 76). Truth is the chief good because (Augustine suggests) nothing could be put above its contemplation.

Truth is also the chief good because, by a quick bit of Scriptural manipulation, Augustine thinks he can show that 'truth itself' is *identical with God*:

'Haec est libertas nostra, cum isti subdimur veritati: et ipse est Deus noster qui nos liberat a morte, id est a conditione peccati. Ipsa enim Veritas etiam homo cum hominibus loquens, ait credentibus sibi: Si manseretis in verbo meo, vere discipuli mei estis, et cognoscetis veritatem, et veritas liberabit vos.' (dLA 2.37)

Hence 'Inmutabile bonum non esse nisi unum verum beatum Deum' (dCD 12.1); and the 'bona voluntas' which we cannot lose without choosing to is only a supremely good thing because it is identical with desire for this God.

This completes my presentation of Augustine's seven definitions of the human good. Perhaps it is clear now that they do in fact tie up, in this way. The highest human good is contemplation (4a) in peace (4b) of God because the beatific vision is the highest good that there is, the highest good that we can desire; such a good is also preeminently the *right* object of desire (4d) and a *righteous* desire (4e) or 'bona voluntas' (4g), and (given that humans can attain to the very heights of the cosmic order) the *natural* thing for humans to desire (4c). Augustine also believes that God is that object of desire which (4f) it is supremely difficult to lose unless you choose to: that is the point of his argument (dLA 1.20-21) that no wicked power in heaven or earth, apart from the defection of the 'voluntas' itself, could be strong enough to overpower the 'bona voluntas'.

The underlying motivation of all these arguments is the same. Augustine's central moral axiom is that there should be a match between the ordering of our desires and choices and the ordering of the world:

'Naturam corporis inferiore gradu esse quam animi naturam, ac per hoc animum maius bonum esse quam corpus.' (dLA 2.48)

'Hic itaque in unoquoque iustitia est, ut oboedienti Deus homini, animus corpori, ratio autem vitiis etiam repugnantibus imperet.' (dCD 19.27)

'Virtutes quas habere sibi videtur per quas imperat corpori

et vitiis, ad quodlibet adipiscendum vel tenendum rettulerit nisi Deum, etiam ipsae vitia sunt potius quam virtutes.' (dCD 19.25)

The perfect ordering of the volitions would reflect perfectly the perfect ordering of the goods which there are in the world; desire would match up perfectly with *desirability* (in the gerundive sense). Thus, for Augustine, ethics depends on the *natural ordering of the world*.

5. Good Will and the Order of the World

And, in turn, the natural ordering of the world depends on God: '[Tu] qui omnia in mensura et numero et pondere disposuisti' (Conf 5.4). As Gosselin remarks (Gosselin 1949, p.525), "'Modus, species, ordo" reviennent comme un refrain': a refrain which crops up, with some variations such as the substitution of 'mensura' for 'modus' and of 'pondus' for 'ordo', throughout Augustine's works.

The point of this refrain is a point about God as the formal cause of that order in the world which is a necessary condition of its continued existence. This is argued at length in dLA 2.44-45.

The first step in the argument is the claim that all non-eternal things have *form* (as opposed to *being form*). They

have this form inasmuch as they admit of numbers, i.e., are measurable:

'Intuere coelum et terram et mare, et quaecumque in eis vel desuper fulgent, vel deorsum repunt vel volant vel natant: formas habent, quia numeros habent: adime illis haec, nihil erunt.' (dLA 2.42)

'Si ergo quidquid mutabile aspexeris, vel sensu corporis vel animi consideratione, capere non potes nisi aliqua numerorum forma teneatur, qua detracta in nihil recidat.' (dLA 2.44)

(Note here that, in both passages, Augustine says that what lacks this form is *nothing*. Likewise, at dNB 3 he writes that 'Ubi nulla sunt [modus, species et ordo], nulla natura est'. Some writers, such as Gilson (*Introduction à l'étude de Saint Augustin*, p.258-259), attempt to distinguish divine *making* and *forming* in Augustine's thought about creation, a 'faire' and a 'parfaire', as if God created matter first and only then imposed form on it, like water poured into a jug or a seal pressed into wax. But, at least from passages like these, it seems rather as if the creative operations of making and forming are, to Augustine's mind, *identical*. For him, what was (*per impossibile?*) only made and not also formed would not be anything actual at all.)

Hence Augustine does not go along with the Manichaeans in saying that matter is essentially evil. For him it is unclear

that matter has *any* essence, evil or not. On the other hand, he does not deny the existence (in some sense) of matter. Rather, what he says about it is that there is an inherent inaccessibility, both perceptual and conceptual, about matter considered in itself:

'Nec ista ergo hyle malum dicenda est, quae non per aliquam speciem sentiri, sed per omnimodam speciei privationem cogitari vix potest.' (dNB 18)

Matter is something mysteriously poised on the brink of being nothing; its only reality is so much *in potentia*, and so little *in actu*, that it is barely anything real at all. There is an important analogue to all this in Augustine's theory of action—as we shall see in the next chapter.

The second step of the argument is the claim that the forms recognisable in mutable things must have been imposed on those things from outside:

'Nulla autem res formare seipsam potest: quia nulla res potest dare sibi quod non habet.' (dLA 2.45)

And where does this imposed form come from? From an eternal form:

'Conficitur itaque, ut et corpus et animus forma quadam incommutabili et semper manente formentur. Cui formae dictum est:

"Mutabis ea et mutabuntur; tu autem idem ipse es, et anni tui non deficient" [Ps.101.27-28].’ (dLA 2.45)

This eternal form is identical with 'providentia', a word which Augustine chooses with an eye on Plotinus' *πρόνοια* (v. *Enneads* 3.2-3):

'Hinc etiam comprehenditur omnia providentia gubernari. Si enim omnia quae sunt, forma penitus subtracta nulla erunt, forma ipsa incommutabilis, per quam mutabilia cuncta subsistunt, ut formarum suarum numeris impleantur et agantur, ipsa est eorum providentia: non enim ista essent, si illa non esset.' (dLA 2.45)

'Providentia', in turn, is identical with wisdom:

'Quicumque iter agit ad sapientiam, sentit sapientiam in via se sibi ostendere hilariter, et in omni providentia occurrere sibi...' (dLA 2.45)

And lastly wisdom is of course identical with God, specifically with the second person of the Trinity:

'Esse patrem sapientiae... aeterno patri sit aequalis quae ab ipso genita est sapientia.' (dLA 2.39)

God, then, is the formal cause of the order of the world. That is part of what it means to talk of him as creator. There is no other possible source for the existence, life and intelligence

manifest in the world (dLA 2.46); and, without his formal causality, Augustine argues not only that there would be no order in the world, but even that there would be no world. (As we have seen, the two propositions are barely separable for Augustine.)

The right ordering of volitions, I said, reflects the right ordering of the world; and we have just established that everything that is right in the ordering of the world is for Augustine a direct consequence of God's formal causality. It follows that all right volition is equally under the influence of God's formal causality. *Whatever is a formal cause of the order of the world, is also a formal cause of the rightly ordered 'voluntas'*. To put the same point a different way: the Good Life is the formal cause of the Good Will. The natural aspiration of human volition, and the right direction for human volition, converge; on the Good Life in Augustine's most adequate characterisation of it, as the beatific vision of God. Gosselin notes (*loc. cit.*) that sometimes Augustine substitutes 'pondus' for 'ordo'; others have noted another equation often made by Augustine, between 'pondus' and 'amor'. There is an upshot to this: that there is an 'ordo amorum'; that the natural and right tendency of the soul or the 'voluntas' is *towards the good*, which is to say, ultimately, *towards God*. In the famous words of *The Confessions*: 'Quia fecisti nos ad te, et inquietum est cor nostrum donec requiescat in te' (Conf 1.1). That, for Augustine, is the ultimate meaning of the directedness or rationality or form of voluntary action: the active soul's natural and (almost) inevitable gravitation is towards its source and final purpose, in God.

This is the conclusion about the directedness and the rationality of the 'voluntas' at which I have been aiming all along in this chapter. However, some doubts may remain. In particular, my conclusion equates the formal causality of the good 'voluntas' with that of the 'voluntas' as such. But to many modern minds there will seem to be an unbridgeable gap between a conclusion about what the 'voluntas' ought to be motivated by, and one about what it is motivated by.

The best way of dealing with such doubts is to turn to a question which may already have seemed pressing anyway: the question 'What about "*mala voluntas*"?'. I attribute to Augustine the claim that it is normally a condition of some action's being voluntary that it should be rational, i.e. directed towards some good objective. Can this claim be convincingly maintained in face of the objection that there is in Augustine plenty of evidence for a doctrine of volition towards *bad* objectives? That is the question of the next, the final, chapter.

Chapter 9

'Mala Voluntas' and the Mystery of Evil

1. Introduction
2. 'Mala Voluntas'
3. An Incomplete Account?
4. A Necessarily Incomplete Account?
5. Manichaeism and 'Privatio Boni'
6. Incompleteness Again
7. Augustine and Voluntarism
8. From Augustine's Rationalism to Augustinian Voluntarism
9. Conclusion

'It is commonly held that Augustine's ontology, in which evil is treated as a privation, does no more than evade the problem which it is professed to solve; and that his deeply Christian sense of the reality of moral evil caused him to relapse into Manichaeism with his doctrine of original sin, in which the Not-Being, the Nothing out of which man was created, is transformed into a Something with fatal power. In fact the originality of Augustine appears just in his steady refusal to hypostatise evil.'

(Burnaby 1938, p.37)

'The first man is summed up in one act: he took the fruit and

ate of it. About that act there is nothing to say; one can only tell it; it happens and henceforth evil has arrived.'

(Ricoeur 1967, p.244)

'The Apostle meets the problem by leaving it unsolved.'

(Waller, 1887)

'In order for an action to be good it must be right in every respect: because good results from a complete cause, while evil results from any single defect, as Dionysius asserts (DN 4).' - Aquinas, ST 2a2ae.110.3

1. Introduction

In Chs.6-7 I argued that Augustine, like Aristotle, teaches that the absence of compulsion and the absence of ignorance are conditions of voluntary action. The conclusion of Ch.8 was that Augustine, again like Aristotle, is also committed to the view that voluntary action is normally, in an important sense, *rational*. The relevant sense is that he believes that voluntary action, to be understood as such, must be directed towards some good. Augustine's doctrine of the rationality of action, then, is a weaker one than Aristotle's, but it is nonetheless a doctrine of the rationality of action.

In Ch.4 I suggested that the three conditions of voluntary action which I had argued for from Aristotle's writings pointed us to a way of dealing with the puzzle about *akrasia* raised in Aristotle's system. To conclude my work on Augustine, I will now suggest that the three parallel conditions of voluntary action

which I have argued for in Augustine suggest an analogous treatment of an analogous problem.

The analogous problem is that posed for Augustine by the admitted existence of what he calls 'mala voluntas'. 'Mala voluntas'- he believes- exists, and like akrasia can be characterised as wilful wrongdoing (deliberate choice of the worse over the better). Because of this belief Augustine, being both a Christian and a classical philosopher, is confronted by a dilemma.

On the one hand: Scripture (e.g. Roms.1.20-21) affirms that there is wilful wrongdoing. Admittedly, in the Bible, the origin of wilful wrongdoing is as mysterious as the origin of suffering. But that it is a reality is strongly affirmed.

On the other hand: against Scripture, classical philosophy from Socrates to the Stoics denies the possibility of wilful wrongdoing, of a knowing, deliberate choice of evil. Even the Plato of the *Republic* presents an account of akrasia, not as a matter of the rational soul's choosing to follow desires it knows to be wrong, but as a matter of its being overpowered by those desires. And, as I have argued in Ch.8, Augustine's philosophical psychology was *not* (pace Dihle) radically different from the kind of accounts which the writers of the classical era had proposed. Like theirs, his took as its most basic principle the axiom that all desire is desire of something intrinsically good. Thus the idea of ἔκων ἀμαρτάνειν was just as problematic for him as for

Socrates (or Aristotle). To offer an account of the psychology of wilful wrongdoing, in the terms of a classical theory of action like this, was simply impossible. Augustine was too good at philosophy to miss this- although (as we shall see) he was also too fond of rhetoric to keep himself from ever indulging in the inconsistency of exploring other, ultimately less fruitful, lines of thought.

So the problem which haunts Augustine's theory of the voluntary is a conflict between the givens of revelation and philosophy. If voluntary action has the inherent directedness towards the good which Augustine, in line with his philosophical tradition, says it has, how can there be even one case of genuine deliberate choice of the worse over the better- as opposed to action which looks like such a choice, but is really ignorant or compelled or irrational action?

It may now be clear why I say this problem is analogous to Aristotle's problem about akrasia. In what follows, I hope to make it clear why I also say that Augustine's treatment of this problem of 'mala voluntas' is analogous to Aristotle's treatment of the problem of akrasia.

2. 'Mala Voluntas'

How can there be 'mala voluntas'? To see Augustine's response to this question, we need to look in detail at how he actually

uses the vital words 'mala voluntas' and associated phrases. The evidence will suggest two points:

1. If the notion of 'mala voluntas' is to be used as an explanatory tool in Augustine's theory of action, then it cannot be used to mean 'a choice of something evil in itself'. What Augustine's use of it marks, is not a belief that his talk of 'mala voluntas' *explains* any choice of something evil in itself, but a belief that at this point in the theory of the voluntary, explanations necessarily run out. For voluntary action is, in Augustine's view, necessarily directed towards the good. Therefore we cannot explain behaviour as voluntary action by supposing it to be directed towards something evil.

2. Hence Augustine's account of the origination of 'mala voluntas' is not only incomplete, but *necessarily* incomplete.

Both points- I shall argue- play vital parts in Augustine's own apologetic strategy. But this may be far from obvious at first sight. I now present evidence against both (1) and (2).

2a. Evidence against (1)

The 'directedness' thesis which I have been attributing to Augustine, the claim that all choice is of something intrinsically good, runs clean against the normal line of exegesis of Augustine, which, on the contrary, credits him with a

'neutrality' thesis about the 'voluntas'. That thesis is that the 'voluntas' is not necessarily directed towards anything. Now Augustine does often talk as if his doctrine of 'voluntas' *does* mean that one can have choices between a good and a bad where it is a matter of 'indifference' (in Hume's sense) which one chooses. Consider these five passages:

(i) 'Motus autem quo huc aut illuc voluntas convertitur, nisi esset voluntarius, neque laudandus cum ad superiora, neque culpandus homo esset cum ad inferiora detorquet *quasi quendam cardinem voluntatis.*' (dLA 3.3)

If the 'voluntas' is a hinge between the good ('superiora') and the bad ('inferiora'), then it may be said that, as a hinge between a door and a door frame is not truly part of either, the point of the image must be that 'voluntas' itself has a *neutral* position between good and bad choices, and is not, by its nature, committed to a general policy of preferring either the good to the bad or *vice versa*.

(ii) '[Deus] non solum magna, sed etiam media et minima bona esse praestitit... Virtutes magna bona sunt... Virtutibus nemo male utitur; ceteris autem bonis, id est mediis et minimis, non solum bene, sed etiam male quisque uti potest... Voluntas ergo quae *medium bonum est...*' (dLA 2.50, 2.52)

The claim that the 'voluntas' is a '*medium bonum*' does not, of course, mean that the 'voluntas' is in the middle, neutral,

between good and evil. 'Media bona' are still 'bona'; they are in the middle between 'magna bona' and 'minima bona', not between 'bona' and 'mala'. The claim is rather that the 'voluntas' is the kind of good which can be used or misused. Its *choices*— it seems— are in some sense neutral between good and evil.

(iii) In the dDA, Augustine insists that our having both good and evil desires is no evidence for the two souls, one good and one evil, which the Manichaeans claimed to discern in the human psychology:

'Cur mihi iam [Manichaei] de ulla re audiendi viderentur? An ut discerem hinc ostendi animarum duo esse genera, quod in deliberando nunc in malam partem, nunc in bonam nutat assensio? Cur non magis hoc signum est unius animae, quae *libera illa voluntate huc et huc ferri, hinc atque hinc referri potest*? Nam mihi cum accidit, unum me esse sentio *utrumque considerantem, alterutrum eligentem*: sed plerumque illud libet, hoc decet, quorum nos *in medio positi fluctuamus*. Ita enim nunc constituti sumus, ut et per carnem voluptate affici, et per spiritum honestate possimus.' (Cp. Conf 8.5)

(iv) 'Interest autem qualis sit voluntas hominis; quia si perversa est, perversos habebit hos motus... omnino pro varietate rerum, quae appetuntur atque fugiuntur, sicut allicitur vel offenditur voluntas hominis, ita in hos vel illos affectus mutatur et vertitur.' (dCD 14.6)

The claim is clearly enough that the 'voluntas' can be *either* 'perversa' or 'recta'. This seems to imply that it is equally likely to be either, which again suggests its neutrality. Augustine seems also to be positing a cause for its turning either way, in the environment in which the 'voluntas' finds itself. Cp. dLA 3.74, where he remarks that 'voluntatem non allicit ad faciendum quodlibet, nisi aliquod visum', and seems to suggest that the causes of sin are, in a sense, in our environment: a claim which I shall consider below.

(v) Most decisively of all, consider dSL 58, where- it might seem- the 'directedness' thesis is explicitly rejected:

'Prius igitur illud dicamus, liberum arbitrium naturaliter attributum [esse] a creatore animae rationali, illa *media vis* est, quae vel intendi ad fidem, vel inclinari ad infidelitatem potest.'

2b. Evidence against (2)

At first sight it seems arguable, too, that Augustine's account of the origination of 'mala voluntas' is not in any important sense incomplete. He seems to have plenty to say to back up the claim that 'mala voluntas' is the origin of evil- under at least five different heads.

(a) On at least one occasion he says that sin, and with it

presumably 'mala voluntas', arises from two sources, one being 'spontanea cogitatio', the other 'persuasio alterius' (dLA 3.29). In either case, of course, it remains true that what is done is 'voluntarium' (loc.cit.). Compare dGCPO 2.42: 'Vitii vero auctor est diaboli decipientis calliditas et hominis consentientis voluntas'.

(b) Sometimes he describes 'mala voluntas' as arising from pride, in particular when he is dealing with the fall of the angels:

'Ille autem angelus magis seipsum quam Deum diligendo subditus ei esse noluit, et intumuit per superbiam, et a summa essentia defecit, et lapsus est.' (dVR 26)

'Non enim ad malum opus perveniretur, nisi praecessisset voluntas mala. Porro malae voluntatis initium quae potuit esse nisi superbia? Initium enim omnis peccati superbia est.' (dCD 14.13)

(c) Sometimes he says that 'mala voluntas' originates in a choice of a *private good* by an 'aversio' from a 'communal' good:

'Veritatem autem atque sapientiam nemo amittit invitus; non enim locis separari ab ea quisquam potest; sed ea quae dicitur a veritate atque sapientia separatio, perversa voluntas est, qua inferiora diliguntur... Habemus igitur qua fruamur omnes aequaliter atque communiter.' (dLA 2.37)

'Perversa enim est celsitudo, deserto eo cui debet animus inhaerere, principio sibi quodam modo fieri atque esse principium.' (dCD 14.13)

This mode of description is readily compatible with the mode which describes 'mala voluntas' as being due to pride, as shown by a passage in the dCD where the two modes are run together. I underline those parts which refer to the 'privacy' mode of description, and italicise those which refer to the 'pride' mode:

'Alii sua potestate potius delectati, velut bonum suum sibi ipsi essent, a superiore communi omnium beatifico bono ad propria defluerunt et habentes *elationis* fastum pro excelsissima aeternitate, *vanitatis* astutiam pro certissima veritate, studia partium pro individua caritate, *superbi*, fallaces, *invidi* effecti sunt.' (dCD 12.1)

(d) Sometimes, as we have seen, he argues that the origin of the 'mala voluntas' is in some sense to do with the *environment* in which the 'voluntas' is placed. dCD 14.6 and dLA 3.7 have been quoted already. There is also 83DQ 40:

[In answer to the question 'Cum animarum natura una est, unde hominum diversae voluntates?':] 'Ex diversis visis diversus appetitus animarum est, ex diverso appetitu diversus adipiscendi successus, ex diverso successu diversa consuetudo, ex diversa consuetudine diversa est voluntas.'

Which brings us to (e): Augustine's very frequent descriptions of 'mala voluntas' as originating in, or a concomitant of, or something connected in one way or another with, 'consuetudo', *habituation* (and/ or 'cupiditas' or 'concupiscentia'). This is, in fact, perhaps Augustine's commonest way of characterising 'mala voluntas', especially in the later works.

The idea is clearly present in the 83DQ passage just quoted, where it is given unusually precise formulation. Here 'visum' leads to 'appetitus', 'appetitus' to 'successus' (cp. NE 1147a35-6), 'successus' to 'consuetudo', and 'consuetudo' to 'voluntas'. Elsewhere, the idea is usually less sharply focused, but often seems to be essentially the same: we have the 'mala' or 'bona' 'voluntas' we have because of our habituation, our constitution, the way we are already.

So in the EEGa1, the stage of unsuccessful struggle against sin, 'sub lege', is that stage where the sinner 'trahitur *pondere temporalis cupiditatis*', and (hence) 'relinquit iustitiam' (EEGa1 46). In the dLA, we read that

'Nec mirandum est quod vel ignorando non habeat arbitrium liberum voluntatis ad eligendum quid recte faciat; vel *resistente carnali consuetudine, quae violentia mortalis successionis quodammodo naturaliter inolevit*, videat quid recte faciendum sit, et velit, nec possit implere.' (dLA 3.52)

Even where good impulses are present, they are smothered by the weight of custom and, the suggestion is, contrary bad impulses.

In the later writings, this kind of description of the 'mala voluntas' as 'voluntas' under the influence of 'cupiditas' or 'concupiscentia', the gravity of original sin which drags us away from the choice of the good, is the subject of whole books (such as *de Gratia Christi et Peccato Originali* Bk.2). As a proponent of the theology of the Fall, it is natural for Augustine to want to argue that, in Adam and Eve, 'mala voluntas' led to 'mala consuetudo'; and that, in us, 'mala consuetudo' leads to 'mala voluntas'. So we read that

'Per arbitrii libertatem factum, ut esset homo cum peccato; sed iam poenalis vitiositas subsecuta ex libertate fecit necessitatem.' (dPIH 9)

And again, that

'Ex vitiis naturae (non ex conditione naturae) [est] quaedam peccandi necessitas.' (dNG 79)

But one begins at times to wonder whether, with reference strictly to fallen humanity, Augustine is arguing that 'concupiscentia' causes 'mala voluntas', or that 'mala voluntas' causes 'concupiscentia'. Or is it that 'concupiscentia' is 'mala voluntas'? Or a mishmash of all three? For example, the

progression from 'consuetudo' to 'voluntas' noted above in 83DQ 40 is actually *reversed* in the *Confessions*, where the subject of the description is not Adam but Augustine:

'Quippe voluntas perversa facta est libido, et dum servitur libidini, facta est consuetudo, et dum consuetudini non resistitur, facta est necessitas.' (Conf 8.5)

So which comes first in fallen humans, 'mala consuetudo' or 'mala voluntas'? In the later works, the rather vague and general upshot of teaching about 'voluntas' and 'consuetudo' tends to be that, whichever way round it is, '[sumus] carnales, infirmi, peccatis obnoxii et ignorantiae tenebris obvoluti' (dCD 10.24). The two states of 'mala consuetudo' and 'mala voluntas' seem to become welded together in Augustine's thinking to the extent that to talk of one is to talk of the other. (Compare the way in which, as noted in Ch.6, ignorance and difficulty became for the later Augustine no longer *exculpating* factors, but the very hallmark of *culpability*.) 'Which comes first?' is treated by the later Augustine as being, at least with reference to fallen humans, a chicken and egg question.

One characteristic argument is that the sinner has freedom from righteousness, the righteous person freedom from sin, but both are, nonetheless, *free*:

'Libertas quidem perit per peccatum, sed illa, quae in paradiso fuit, habendi plenam cum immortalitate iustitiam...

liberum arbitrium usque adeo in peccatore non perit, ut per ipsum peccent maxime omnes qui cum delectatione peccant et amore peccati et hoc eis placet quod eos libet.' (cDEP 1.5; cp. dLA 1.37)

This passage seems to present another version of the old chestnut about 'non posse peccare'/'posse non peccare'- a distinction which I have already considered in Ch.7, §4.

'Mala voluntas' is not inconsistent with 'libera voluntas'; but the range of options for a 'mala voluntas' is limited, and different from that available for a 'bona voluntas'. 'Mala voluntas', one might almost say, *becomes* 'peccatum originale':

'[Peccatum] latet, donec repugnans iustitiae malum eius prohibitione sentiatur, cum aliud iubetur atque adprobatur, aliud delectat atque dominatur.' (cDEP 1.17)

3. An Incomplete Account?

I have amassed a good deal of evidence against my two leading claims (1) and (2) about the 'mala voluntas'. How then am I to defend those claims? I will begin by treating of the evidence against (2). This treatment, I hope, will show how my defence of (1) is going to go.

(2) was the claim that Augustine's account of the origination

of 'mala voluntas' is necessarily incomplete. This has suggested the retort that, on the contrary, Augustine's account of the origination of 'mala voluntas' is full, detailed, and therefore perfectly adequate.

Augustine's account (or rather 'accounts'), as given above, may well be full and detailed. But nothing he says under the five heads considered above deals adequately with a central problem which confronts his doctrine of 'mala voluntas'. There is a serious lacuna in his teaching as so far expounded: one which can only be remedied by taking (with Augustine himself, in many places) a view of the origination of 'mala voluntas' radically different from all the above five. This lacuna, I will now argue, seriously compromises any attempt to extract from Augustine's writings a consistent doctrine of the will in the sense required, for example, by Albrecht Dihle as quoted in the epigraph to this chapter. What is this lacuna?

It is one which has already been observed by at least one sharp eyed Augustine scholar, Robert Brown, who (in a seemingly neglected article in *The Journal of the American Academy of Religion*, 1978) anticipates quite a lot of the argumentative strategy I intend to employ here. Let us begin with Brown's response to the suggestion that 'Pride is the cause of the fall', a suggestion in effect equivalent to (b) above:

'What a grand solution this is: Satan turned away from God because he was proud, and thereby became a self centred rebel.

But just a minute! Why was he proud? Did God create him proud? Certainly not, for then God would be responsible for his fall. Did he make himself proud by the free exercise of his will? So he must have done, if Augustine's intent to defend the initial freedom and responsibility of the will is to remain intact. If Satan made himself proud, then this act of will is itself the fall, and not a "cause" of his falling. Pointing to pride therefore cannot constitute an explanation for the fall (an account of *why* the first evil will willed as it did). It is only the substitution of a synonym for the inexplicable free act of falling itself. This substitution of "becoming proud" for "falling" or "first willing evil" is attractive because, by drawing an analogy to the everyday human sin of pride, it makes Satan's act more vivid... but it explains nothing, it in no way renders Satan's fall understandable.'

(Brown, 1978)

Brown's point applies in an obvious way to what I am considering, the question of what Augustine can tell us about the origination of 'mala voluntas'. It won't do to say (b) that pride is the source of the 'mala voluntas': for unless there is *already* a 'mala voluntas', how can there be any (culpable) pride? And this point, about what a follower of Ricoeur might call the 'already'-ness of sin, can be deployed against the other four explanations of the origin of 'mala voluntas' offered in response to the allegation that Augustine's account is necessarily incomplete.

(a) Action which counts as a case of sin, and so of 'mala voluntas', may well originate either through 'spontanea cogitatio' or 'persuasio alterius'. But either way, it will *not be sin* (or 'mala voluntas') unless the 'cogitatio' or the 'persuasio' was *not* a sufficient and necessary condition of the action's occurring. For if either *was* such a condition, then the agent was *not* free to do otherwise than she did, and so was not responsible. Thus, if (a) is an explanation of the origination of sin, in the sense of wilful wrongdoing, it cannot be a *complete* explanation; but, if (a) *is* a complete explanation, it cannot be an explanation of *sin*.

(c) To say that 'mala voluntas' originates in the turning away from the shared good to a private good simply prompts the question whether this turning away is *itself* a bad action. If it is *not* a bad action, then why does 'mala voluntas' result from it? If it *is* a bad action, then mustn't 'mala voluntas' be present *already* when it is performed- in which case the origin of 'mala voluntas' has not been given? The idea that 'mala voluntas' can be *characterised* as 'aversio' from the 'commune bonum' (or (b) as pride) seems all right. The idea that it can be (efficient causally) *explained* in these ways does not.

(d) As for (a): action which counts as a case of sin, and so of 'mala voluntas', may well originate through the influence of factors in the environment which are perceived. But if so, it will *not be sin* (or 'mala voluntas') unless the environmental influence in question was *not* a sufficient and necessary

condition of the action's occurring. For if that influence *was* such a condition, then the agent was not free to do otherwise than she did, and so was not responsible. Thus, if (d) is an explanation of the origination of sin, it cannot be a *complete* explanation; but, if (d) *is* a complete explanation, it cannot be an explanation of *sin*.

(e) is the most interesting of the five accounts; but (as noted) it is very unclear precisely what this account comes to. If, on the one hand, the point is either that 'concupiscentia' originates in 'mala voluntas', or that the two are equivalent conditions, then it is patent that the origin of 'mala voluntas' has not been explained. If, alternatively, the point is that 'mala voluntas' originates in 'concupiscentia', then the same problem as for (b) and (c) recurs. Is this 'concupiscentia' *itself* bad? If it is *not*, then why does 'mala voluntas' result from it? If it *is*, then mustn't 'mala voluntas' be present *already* when the 'concupiscentia' arises? Again, in either case, the origin of 'mala voluntas' has not been explained.

I take it that these arguments vindicate at least the claim that Augustine's account of the origination of 'mala voluntas' is incomplete— if not my stronger claim (2) that Augustine's account of the origination of 'mala voluntas' is *necessarily* incomplete. The lacuna I have indicated is simply this: that not one of these supposed accounts of the origination of 'mala voluntas', *is* such an account. None of them delivers the goods required by actually telling us where 'mala voluntas' comes from. And this does seem

to suggest that there is something incomplete about Augustine's account.

4. A Necessarily Incomplete Account?

What more is needed, then, for me to move from saying that that account is incomplete to saying that it is *necessarily* incomplete? The missing steps- as I shall now argue- are provided by Augustine himself. Augustine himself gives us (the makings of) a valid argument to the conclusion that such attempts at explaining the origination of 'mala voluntas' as the above five *cannot*, in principle, succeed; and indeed, that no attempt whatever to explain its origination could succeed. This argument which I am going to attribute to Augustine will explain, with Brown, why 'the first evil will must be inexplicable'. But the argument I propose will give us more reason for saying that than Brown's argument does. It will also lead us from the establishment of the second claim made at the start of this chapter (the incompleteness thesis) to the establishment of the first claim made there (the directedness thesis).

That there is this argument in Augustine will allow us to see (a)-(e) in a rather different light. On the best interpretation of his work, Augustine is not offering us any kind of positive explanation of the origin of 'mala voluntas' at all. Rather, what he means to offer us is an explanation of why there can be no such explanation. Hence it will emerge that neither (a)-(e), nor

anything else in Augustine's writings, count as possible ingredients for a 'theory of will' which is to explain the origin of evil by reference to the neutrality, relative to good and evil, of the will's power to choose. One might, indeed, take the failure of (a)-(e) to explain the origin of 'mala voluntas' as a kind of *reductio ad absurdum*: a demonstration of the hopeless task which any similar attempt at explanation of 'mala voluntas' is, necessarily, setting itself.

For, sometimes almost in the same breath as he offers us one of (a)-(e)- or some other attempt at explaining 'mala voluntas'- Augustine also makes remarks like these:

'Improba voluntas malorum omnium causa est... Tu autem si huius radicis causam requiris, quomodo erit ista radix omnium malorum? Illa enim erit quae causa huius est, quam cum inveneris, etiam ipsius causam quaesiturus es, et quaerendi nullum habebis modum. *Sed quae tandem esse poterit ante voluntatem causa voluntatis?*' (dLA 3.49)

Augustine says quite clearly here that it is impossible to explain the origination of 'mala voluntas'. But why should that be? Because 'what is not anything, cannot be known' (dLA 2.54). Is Augustine, then, saying that the origin of 'mala voluntas' is *nothing*, or nothingness?

It seems clear that Augustine is indeed saying something like that. This is a recurring idea in his thought. The claim that

there is some sort of connection between wickedness and nothingness is first hinted at in the dBV:

'Etenim ipsam nequitiam matrem omnium vitiorum, ex eo quod nequidnam sit, id est ex eo quod nihil sit.' (dBV 8; cp. Cicero, *Tusculanae Disputationes*, 3.18)

In the dDA, likewise, the origins of evil in general are said to be, at best, highly mysterious. The Manichaeans' urge to look into the questions of pain and sin is an urge '*res abditas intueri*' (dDA 9). The Manichaeans' attempts to make problems for Catholics by raising unanswerable questions about sin and pain are disingenuous: 'Si ab homine, unde homo? Si ab angelo, unde angelus? ...quasi per quamdam catenam ad Deum mala connecti. Hac quaestione illi regnare se putant, quasi vero interrogare sit scire'. Augustine's wry comment is significant: '-Utinam id esset! nemo me scientior reperitur' (dDA 9). Against the Manichaeans, Augustine is arguing that these are questions without answers.

In the *Confessions*, Augustine's long meditation on the stealing of the pears (Conf 2.4-10) gives us another question which remains unanswered: *Why* did Augustine steal the pears? He says he got a 'kick' out of stealing the pears- *because* it was forbidden (2.4). Was *this* his reason for stealing them? But such a motive seems hardly an adequate reason for doing anything! Yet why else should Augustine have done it? For the taste of the pears (2.4), for their fairness (2.6), for the company of the

'adulescentuli' with whom he did the deed (2.8), for the joke of the thing (2.9)? None of these answers seems to satisfy Augustine. The action remains puzzling precisely because it was so random, so aimless, so *lacking* in motivation. Once again, this suggests to him the link between wickedness and nothingness, the idea that wickedness is mysterious *because* of its essential nothingness:

'Dicat tibi nunc ecce cor meum, quid ibi quaerebat, ut essem *gratis* malus, et malitiae meae *causa nulla* esset nisi malitia.'

(Conf 2.4)

'Quis exaperit istam tortuosissimam et implicatissimam nodositatem? Foeda est; nolo in eam intendere, nolo eam videre.'

(Conf 2.10)

But what does it mean to talk like this? What is gained by saying that the origin of 'mala voluntas' is nothing, or nothingness? I believe that the answer to this can be given by reminding ourselves of what was established in Ch.8 as a fundamental principle of Augustine's philosophy of action: the 'directedness' thesis.

The 'directedness' thesis, we saw, is the claim that all voluntary action is *rational* in the sense of being necessarily directed towards some good or other. To put it another way: to *explain* an agent A's piece of behaviour B as a voluntary action is (1) to posit a good G at which B could have been directed by

A, and (2) to surmise that B was directed at G by A.

Now, how might such an account apply to the kind of action which is typical of 'mala voluntas': a deliberately wicked action, a piece of wilful wrongdoing? The suggestion being canvassed here (in the form of claim (1)) is that it is precisely in its application to wilful wrongdoing that such an account must necessarily break down. We may make a sharp distinction between the case in which someone does what is, let us say, 'objectively' wrong, but where that action is nonetheless the best action available to him (simple wrongdoing); and the case where someone does what is 'objectively' wrong *even though he knew it was wrong and could have avoided it* (wilful wrongdoing).

Simple wrongdoing, so defined, will be perfectly easily explicable: the agent had no choice, or the agent at least did the best he could. But *wilful* wrongdoing, so defined, will not only be inexplicable: it will be *necessarily* inexplicable. The agent was not compelled; the agent was not ignorant of what he was doing; and the agent did not hold that what he did was *really* the right thing to do. And yet he did it. Such an action is, by its very nature, inexplicable, simply because the explicability of an action can only mean the possibility of relating it to some good at which it is supposed to aim. But if *no good whatever* is aimed at by an action as correctly described, then of course we cannot specify any good to which it is related; and hence it is necessarily true that we cannot explain it. Our ambitions to explain human actions, and to see them as cases of *wilful*

wrongdoing, are ambitions which necessarily, in the nature of the case, pull against each other.

This, I believe, is an argument which Augustine does actually make out. Formally, its premisses are (A) that explanation of an action can only mean stating what good it aims at, and (B) that 'mala volūntas', 'wilful wrongdoing' in the strictest sense, means action which aims at no good whatever. From these it validly derives the conclusion (C) that wilful wrongdoing in this sense is necessarily inexplicable. Direct exegetical evidence can be found for all three steps of the argument. Plenty of evidence has already been given for (A) in Ch.8. I here present the evidence for (C) and (B); in that order, since (B) is the more complicated of the two.

(C) 'Nemo ergo ex me scire quaerat, quod me nescire scio: nisi forte *ut nescire discat, quod sciri non posse sciendum est.*' (dCD 12.7)

'Unde igitur erit [motus peccandi]? Ita quaerenti tibi, si respondeam nescire me, fortasse eris tristior: sed tamen vera responderim. *Sciri enim non potest quod nihil est...* Quoniam defectivus motus est [aversionis], omnis autem defectus ex nihilo est, vide quo pertineat.' (dLA 2.54)

'Quid opus est quaerere unde iste motus existat?' (dLA 3.2)

(B) At dLA 2.54, Augustine tells us that '*defectivus motus*

est [aversionis]'. What does this mean? One way of clarifying it is to compare dCD 12.7:

'Nemo igitur quaerat efficientem causam malae voluntatis: non enim est efficiens, sed *deficiens*, quia nec illa effectio sed defectio. Deficere namque ab eo, hoc est incipere habere voluntatem malam. Causas porro defectionum istarum... velle invenire tale est, ac si quisquam velit videre tenebras vel audire silentium, quod tamen utrumque nobis notum est... non sane in specie, sed in speciei privatione.' (dCD 12.7)

And, again, the analogies which Augustine offers here are worth considering. There is no sight at all without light rays; no hearing at all without sound waves (and the Psalm says: 'In thy light we see light'). The absence of either is something which can only be understood relative to the *presence* of either. The Manichaeans' mistake- their dualism of equal and opposite good and evil- is like thinking that darkness is a colour; it is to think that evil is one of the things which can be explained, when in fact its very presence is the *negation* of explanation. But what is it which is absent, in the case of an act of wilful wrongdoing, which is present in some other case and there makes explanation possible?

Note that Augustine speaks here of the 'causa deficiens' of wilful wrongdoing, in sharp contrast with the 'causa efficiens' which some other things have. What other things? I suggest, in line with (A), that the contrast is between actions which are

cases of wilful wrongdoing and actions which are not. Then actions other than wilful wrongdoing *will* have 'causae efficientes'. An action which has a 'causa efficiens' will be a normal, non-skew, explicable case of action: either a good action, or a piece of non-wilful wrongdoing. The case of wilful wrongdoing, where there is not so much (as Augustine's pun has it) a 'causa efficiens' as a 'causa deficiens', will be the queer and inexplicable case.

This will sound somewhat odd if we translate 'causa efficiens' in the obvious way, as 'efficient cause'. Does Augustine really mean to talk about efficient causation in this context? Or is his choice of this term dictated not so much by philosophical considerations as by a rhetorical (and pun-making) impulse, by his playing-off of 'deficiens' and 'defectio' and (if I am right to bring in dLA 2.54) 'defectivus'? The way *this* passage (at any rate) continues suggests that, if he were to have made this distinction, he would have said that here he was talking about a presence (or absence) of *final* causality, not of *efficient*. The defection from God is not here seen as underdetermined by efficient causality, so that it would be random. Here it is underdetermined by *final* causality, so that it is *inexplicable*.

In other places, however, there is clear evidence that Augustine thought that talk about the causes of 'mala voluntas' did not exclude talk about efficient causes. For example:

'Improba voluntas malorum omnium causa est... Tu autem si huius radicis causam requiris, quomodo erit ista radix omnium malorum? Illa enim erit quae causa huius est, quam cum inveneris, etiam ipsius causam quaesiturus es, et quaerendi nullum habebis modum. Sed quae tandem esse poterit ante voluntatem causa voluntatis?' (dLA 3.48-49)

The point here seems to be that, while a 'mala voluntas' exercises efficient causality over other states of affairs in the world, nothing else can exercise efficient causality over *it*. For if it were so, it would not be able to do otherwise, and so not responsible (cp. my arguments v. (a), (d) and (e) above). Likewise, we read that

'Causa miseriae malorum angelorum [est] quod ab illo qui summe est, aversi ad se ipsos conversi sunt, qui non summe sunt... Huius porro malae voluntatis causa si quaeritur, nihil invenitur. Quid est enim quod facit voluntatem malam, cum ipsa faciat opus malum? Ac per hoc mala voluntas efficiens est operis mali, malae autem voluntatis efficiens nihil est. Quoniam si res aliqua est, aut habet aut non habet aliquam voluntatem; si habet, aut bonam profecto habet aut malam. Si bonam, quis ita desipiat, ut dicat quod bona voluntas faciat voluntatem malam?' (dCD 12.6)

The angels' misery is caused (this is efficient causality) by their 'mala voluntas'; but their 'mala voluntas' is not similarly caused by anything, for if it were it would not be culpable-which (Augustine has already decided) it is.

Thus, quite possibly, Augustine would hold that in a universe which begins perfect, any deficiency of formal or final cause would have to be mirrored by a deficiency of efficient cause. That, however, is as may be. My main concern here is with final (and formal) causality. The evidence that may be used for (C) may also suggest a rather different argument, about efficient causality, to the argument (A)(B)(C), about final causality. The point here is that that evidence does in any case support the final-causal argument.

If this evidence adds up, then it shows that Augustine does indeed teach that actions which are wilful wrongdoings are necessarily inexplicable, because to explain a voluntary action is to indicate the good at which it aims, and a wilful wrongdoing aims at no good. And this in turn shows what I meant by claim (2): that his account of the origin of 'mala voluntas' is not only incomplete, but necessarily incomplete. For (claim (1)): necessarily, there can be no explanation of what is *inexplicable*. And this is (or ought to be) Augustine's final line of defence against all comers on the question 'Unde malum?'

Thus there is a fundamental distinction in Augustine's thought, not always as carefully drawn by the saint himself as one might have liked, between what I have called 'simple wrongdoing' and 'wilful wrongdoing'. There is no mystery about simple wrongdoing; it is explained by reference to the good at which the agent is (mistakenly) aiming. But there is a mystery about 'wilful wrongdoing': because the action in question is not

aimed at *any* good, mistakenly or otherwise.

5. Manichaeism and 'Privatio Boni'

To illustrate this remark: consider Augustine's response to the Manichaeism teaching about the 'two souls', in his important anti-Manichaeism pamphlet *de Duabus Animabus* (dDA). In line with their dualist metaphysics, the Manichaeism argued for a morally dualist psychology. On their view, just as evil was simply the mirror image of good, so evil desire was simply the mirror image of good desire. The existence of such radically opposed desires in each person was taken to be evidence for the existence of two souls in each person. The Manichaeism's good soul is *perfectly* good, and their bad soul is *perfectly* bad: the good soul, which is part of the 'ipsa substantia Dei' (dDA 16), pursues the greatest available good by the shortest available route. The bad soul, which is absolutely nothing to do with the good God ('nulla prorsus ex parte ad Deum [pertinet]', dDA 16), pursues the greatest available evil with equal assiduity. It is as if a mediocre chess player's mediocrity were due to the evenly balanced influence on her of the ghosts of two Grand Masters haunting her soul. Both urge moves on her; but one is playing straight chess, the other suicide chess.

In metaphysics, the Manichaeism were centrally committed to this claim:

(M) Just as there are substances whose very nature is to be good, so equally there are substances whose very nature is to be evil.

In moral psychology, they were centrally committed to this claim:

(P) Explanations of action can just as well have the form: 'A does B because A thinks it will bring about C, and A thinks C is a good result' as the form: 'A does B because A thinks it will bring about C, and A thinks C is a *bad* result'.

Now Augustine believes that both these claims are incoherent, and for related reasons. (M) He believes that the idea of an evil substance is a contradiction in terms. Existence as such is, in his view, a good (dNB 17, dVR 44); goodness pertains to substances naturally, since they were all made by God (dNB 12): 'Omnis ergo natura bona est' (dNB 3; dLA 3.22).

Therefore there is no symmetry between the good and the bad; the mirror image idea is exactly wrong. For a substance to be bad is not for it to partake of an equal and opposite quality to goodness: it is for it to be, somehow, a privation of good ('privatio boni'), *less than it naturally should be* (dNB 17, Ench 11, dVR 44). As Augustine remarks at dNB 6, any nature which can be corrupted must have some good in it.

The existence of evil is parasitic upon the existence of

good. Evil is not so much like a bad quality or thing, as like the absence of a good quality or thing (dNB 17). Different evils have no independent natures of their own, as good things have (dDA 1). They take their natures entirely from the good things of which they are 'privations' (dDA 6), just as the sense of any negative proposition depends not so much on the negation itself as on the affirmative proposition which it negates (dDA 6). To assert that 'x exists', where x is evil, is itself to assert something which (if true) detracts from x's evil. Hence there can exist no perfect evil except non-existence; which (Augustine thinks) means the same as 'Perfect evil does not exist' (dNB 9, 17).

Here is the relevance of the 'privatio boni' argument to the 'two souls' hypothesis. All substances, as such, are good (dNB 17); the soul is a substance (dDA 7); so all souls, as such, are good (dDA 1); so there is no evil of souls symmetrical in kind to the goodness of souls.

'Est autem vitium primum animae rationalis voluntas ea faciendi quae vetat summa et intima veritas. Ita homo de paradiso in hoc saeculum expulsus est, id est ab aeternis ad temporalia, a copiosis ad egena, a firmitate ad infirma; non ergo a bono substantiali ad malum substantiale, quia nulla substantia malum est; sed a bono aeterno ad bonum temporale, a bono spirituali ad bonum carnale, a bono intelligibili ad bonum sensibile, a bono summo ad bonum infimum. Est igitur quoddam bonum, quod si diligit anima rationalis, peccat, quia infra illam ordinatum est. Quare

ipsum peccatum malum est, non ea substantia quae peccando diligitur.' (DVR 38)

(P) Likewise, as has already been pointed out, the idea of explaining an action with reference to the *bad* aim which it pursues is also, for Augustine, a contradiction in terms. Wilful wrongdoing, 'mala voluntas', cannot be explained symmetrically to 'bona voluntas', deliberate choice of the good. For it is for Augustine something very like a conceptual truth that explanation of any voluntary action means relating it to a desired *good*. To perceive something as a bad is, necessarily, to be disposed to avoid or avert it, not to pursue it. Someone who (like the Manichees) does not believe this has simply misunderstood, or misused, the word 'bad'.

Augustine's response to the Manichaeans is to reject their two symmetries. If there *were* such symmetries, then Augustine would indeed be forced to admit that something like the Manichaean 'two souls' hypothesis is true. But, Augustine argues, there is no such symmetry between good and evil in psychological explanations, any more than there is in metaphysics. There are, rather, two kinds of bad action: wilful wrongdoing, where no good is chosen, which (as we have seen) is not explicable at all; and simple wrongdoing, where the wrong good is chosen through ignorance or compulsion (culpable or otherwise), which is explicable as a degenerate case of good action.

'Quare non duas animas hinc fateri cogor? Possumus enim

melius et multo expeditius intelligere *duo genera bonarum rerum*, quorum tamen neutrum ab auctore Deo sit alienum, *unam animam* ex diversis afficere partibus.' (dDA 19)

Augustine's argument is that wrongdoing, at least of the simple variety, does not display desire of any evil, but desire of some inappropriate kind of good. His examples of goods which are often inappropriate tend to be physical goods, 'sensibilia', as opposed (usually) to 'intelligibilia'. While both 'sensibilia' and 'intelligibilia' are repeatedly affirmed to be good, still 'intelligibilia' are far better than 'sensibilia':

'Ut cuncta quae tactu et visu, vel quolibet alio modo corporaliter sentirentur, tanto essent inferiora his quae intelligendo assequeremur, quanto ipsos sensus intelligentiae cedere videremus.' (dDA 2)

The pursuit of intelligible goods must take precedence in any well ordered life over that of sensible goods. As I suggested in the last chapter, the right ordering of one's desires reflects the real ordering of the cosmos: good desire has as its formal cause Reality with a capital r, which is ultimately identifiable with God Himself (dLA 1.16-20; cp. dNB 1,3, dDA 2-4).

As justice, for Plato, is 'giving each his due', so the just life, for Augustine, is giving to each kind of good the attention which its place in the order of goods makes proper. And the unjust life, correspondingly, is failing to do so (dLA 2.48-52).

No object of choice is bad, and no act of choice is bad, except in the sense in which it could have been better (dVR 78). 'Peccatum non est appetitio naturarum malarum sed desertio meliorum' (dNB 34). In the dDA, the image Augustine uses for this 'desertio' is that of 'imitatio'; the image also appears in the *Soliloquies* (2.11, on falsity and resemblance), and in the dVR (66). The point stressed is that 'imitatio' is not bad because of the intrinsic badness of the models which it takes, but because of the unfittingness of the 'imitatio' to the nature of the creature guilty thereof.

'Peccando fiunt malae non quia malas, sed quia male imitantur... Recte igitur plerumque improbamus imitantem, cum autem quem imitatur probemus. Improbamus autem, non quia non sit assecutus, sed quia omnino assequi voluit... Lucens luna laudatur, suoque cursu placet; tamen si eam sol velit imitari, cui non summe ac iure displiceat?' (dDA 20)

6. Incompleteness Again

It may seem that this account works well enough to describe 'simple' wrongdoing, but does not apply to *wilful* wrongdoing. The above account, emphasising against the Manichaeans the derivative nature of sin, applies nicely to those who are already in the bondage of sin, and cannot help but follow their wicked propensities. What it completely fails to explain is how anyone

not already in that state could get there in the first place.

'Fieri enim potest ut propria [animae] illae voluntate appetendo quod non licebat, hoc est, peccando, ex bonis factae sint malae.' (dDA 20)

Augustine's words are 'fieri potest'; but what he does not give us, here or anywhere else, is any indication of *how* this can happen. The notion of simple wrongdoing is a clear and comprehensible idea. But the existence of simple wrongdoing, it turns out, must always be causally dependent on the prior existence of wilful wrongdoing. And this concept is radically incomprehensible. Which shows, again, the necessary incompleteness of Augustine's account.

The account, in fact, yields a regress, as is hinted at dCD 12.6:

'Quid est enim quod facit voluntatem malam, cum ipsa faciat opus malum? ...Si res aliqua est, aut habet aut non habet aliquam voluntatem; si habet, aut bonam profecto habet aut malam. Si bonam, quis ita desipiat, ut dicat quod bona voluntas faciat voluntatem malam?' (dCD 12.6; cp. dLA 3.49)

If simple wrongdoing is failure to act on the *right* or *appropriate* good desire, this failure to act on the right good desire must itself have been either a voluntary choice to fail, or an involuntary failure. If it was an involuntary failure, then

the agent could not help acting on the wrong good desire, and so was not culpable for her failure. If it was a voluntary choice to fail, then the agent could help it; and so was culpable for her choice to fail to act on the right good desire. But if the agent was culpable for this choice, then this choice must have been a (simple?) wrongdoing. And the definition of simple wrongdoing is 'an agent's failure to act on the right good desire'. So the agent's choice to fail to act on the right good desire was itself due to *another* failure to act on the right good desire. And again: this failure was either a voluntary choice to fail, or an involuntary failure...

To put it another way. Augustine says explicitly that we should explain wrongdoing by pointing out the good at which its agent aimed:

'Cum itaque de facinore quaeritur, qua causa factum sit, credi non solet, nisi cum appetitus adipiscendi alicuius illorum bonorum, quae infima diximus, esse potuisse adparuerit, aut metus amittendi.' (Conf 2.5)

But if I do so explain someone's wrongdoing, then (on Augustine's theory) I have not yet completely explained his motivation. For in doing the wrong act, he (*ex hypothesi*) voluntarily aimed at a lesser good than he might have done: that was why it was wrongdoing. But why did he so aim? To explain this is to specify the good at which he aimed in so lowering his sights. But if he thought there was a good to be achieved by

lowering his sights in this way, nonetheless, if his action was wrongdoing, he must have been able to choose a better good: so why didn't he? *Et cetera*.

This regress appears because of Augustine's refusal to allow the Manichaeans their symmetry between pursuit of good and bad objectives. Wherever it is possible to explain wrongdoing, we must necessarily explain it as *simple* wrongdoing. Now such explanation seems to depend upon a prior notion of *wilful* wrongdoing. But, unlike simple wrongdoing, wilful wrongdoing is not explicable in terms of a pursued good. But that does not mean it *is* explicable in terms of a pursued evil; for there is no such type of explanation. What it means is that wilful wrongdoing is not explicable at all. And here, once again, the necessary incompleteness of Augustine's account of 'mala voluntas' is obvious. My suggestion is that this kind of necessarily incomplete account is no mishap; it is what Augustine is *aiming at*.

7. Augustine and Voluntarism

I conclude that, for Augustine, wilful wrongdoing simply cannot be explained: it is a mystery. In the case of a piece of wilful wrongdoing, the agent sees and acknowledges what it is right to do, but, somehow, inexplicably, *reneges* on this knowledge. That Augustine considers this kind of renegeation *inexplicable* is his concession to the claims of classical

philosophy; that he admits that it is *possible* is his concession to the doctrine of Scripture.

That, of course, was the dilemma I mentioned at the outset. Revelation says that there is wilful wrongdoing; philosophy says that there can't be. Augustine's most adequate view of the origin of wilful wrongdoing rather cleverly combines elements of both sides. First, analogously to Aristotle's treatment of the problem of akrasia, Augustine argues that wilful wrongdoing is rationally inexplicable in the sense explored above. Second, however, he insists that wilful wrongdoing does occur— the evidence being in the revelation of Scripture. The fact of wilful wrongdoing is a revelation to faith; to intelligence, it is a *mystery*. (There is even a biblical phrase which, tenuously, Augustine might have cited in support of this theory: τὸ μυστήριον τῆς ἀνομίας (2 Thes.2.7). But for the evidence that he did not take this opportunity, see dCD 20.19.)

So, besides the analogies which I have suggested, there is also an important contrast between Aristotle's and Augustine's theories about wilful wrongdoing. Aristotle's theory (as I argued in Ch.4) suggests the conclusion that the existence of full akrasia would be a *logical impossibility*. Augustine's theory, on the other hand, suggests the conclusion that the existence of 'mala voluntas' is *logically inexplicable*. Aristotle's position is that 'There *can't* be wilful wrongdoing; therefore there *isn't*'. Augustine's (best) position is that 'Wilful wrongdoing *makes no kind of sense*; yet *there it is*'. In this contrast is the

biggest difference between Aristotle's and Augustine's theories of action. It is the point at which they decisively part company; and it is the point at which one should acknowledge that, if what I called Aristotle's rationality condition of voluntary action does *usually* have an analogue in the conditions of voluntary action which Augustine assumes, there is a crucial exception to this rule. For Augustine thinks that wilful wrongdoing is, in the sense I have explored, 'irrational'; but this does not lead him to deny that it is *voluntary*.

It is also the point at which the commentators start talking about Augustine's alleged belief in 'voluntarism', or 'irrationalism', or in a 'theory of the will'. So Brown, in the article already quoted, undertakes to focus on

'...that primal act in which Adam freely turned his (*wholly undetermined*) "voluntas" away from God...'

(Brown 1978, p.318: emphasis added).

So also Dihle writes:

'The key role attributed to will in Saint Augustine's corresponding systems of psychology and theology resulted mainly from self-examination. It is not derived from earlier doctrines in the field of philosophical psychology or anthropology, and seems to mark a turning point in the history of theological reasoning. From Saint Augustine's reflections emerged the concept

of a human will, prior to and independent of the act of intellectual cognition, yet fundamentally different from sensual and irrational emotion... It is mainly through this entirely new concept of his own self that Saint Augustine superseded the conceptual system of Graeco-Roman culture.'

(Dihle 1982, p.127)

What I want to argue, in the remainder of this chapter, is that it is now evident why this standard view gets Augustine wrong. He is not offering us a 'theory of the will' of the kind supposed by Brown 1978, Dihle 1982, MacIntyre 1985, Thonnard 1941 *et alii*. If he has a 'theory of the will' at all, it is about '*the will*' in a much thinner and less general sense than these commentators suppose. The key role, in his understanding of wilful wrongdoing, is not played by a concept of *will* at all, but by a concept of *mystery*.

Contrast these two claims, a specific one and a general one:

X. For Augustine, the human agent is- somehow- mysteriously capable of wilful wrongdoing which is irrational action. In such action, *and nowhere else*, a voluntary choice is made which has no rational justification even from the agent's own viewpoint.

Y. For Augustine, voluntary action *in general* either has not, or does not need, any kind of rational justification.

I believe that (X) is perfectly correct. But (X) is not consistent with (Y), which I take it is the claim central to the 'voluntarist' view of Augustine's theory of action. Nor can (Y) itself be sustained as an interpretation of Augustine.

My reasons for saying this should already be clear. In the first place, I have already argued at length- the claim was central to the last chapter- that Augustine is committed to the theses (i) that voluntary action is, or is usually, rational, and (ii) that voluntary action is rational, i.e, explicable, only insofar as it is directed towards something good. Therefore (in the second place), it is simply untrue that either *good* action, or *simple* wrongdoing, is lacking in rational justification. And hence (thirdly), it is emphatically *not* the case that Augustine believes that voluntary action *in general* either has not, or does not need, any kind of rational justification.

True, he certainly believes that *wilful wrongdoing* has no kind of rational justification. (Indeed, if we take the efficient causal interpretation of dLA 3.49 and dCD 12.6 suggested above, he even believes that wilful wrongdoing has no sufficient reason for its occurring.) But the whole point of this as a hypothesis about wilful wrongdoing is that wilful wrongdoing is *the exceptional* case. What makes it so mysterious is exactly that it breaks an *otherwise unexcepted* rule: the rule that all voluntary action, by nature, necessarily has some kind of rational justification.

It may now be evident how I can answer one important remaining question. In §2, I amassed a fair amount of evidence for the contradiction of the 'directedness' thesis: the thesis of the *neutrality* of the 'voluntas'. How can I account for this evidence? Do I simply have to dismiss it as an inconvenient inconsistency? I think not. I take it that these passages are affirmations, not of (Y), but of (X).

That is: Augustine is affirming, in these words, the brute fact of our ability to renege on what we recognise as the right course of action. Even at his most emphatic in these passages, Augustine says no more than the 'voluntas' 'illa media vis est, quae vel intendi ad fidem, vel inclinari ad infidelitatem potest' (dSL 58). Need even this affirmation imply any more than Augustine's acceptance of the paradox I have already attributed to him- that 'mala voluntas' is inexplicable, and yet it happens? I suggest not: I suggest that there is, in fact, nothing in these passages that shows Augustine supporting any stronger thesis than (X). Hence, they are no evidence for the claim that he holds to (Y).

The traditional interpretation of Augustine sees him not as a 'rationalist' (i.e., one who insists that, as such, voluntary action should normally be in some sense rational) but as a 'voluntarist' (i.e., a subscriber to (Y)). But I believe that those who follow the traditional interpretation are standing Augustine on his head. The tradition has it that Augustine is a thoroughgoing 'voluntarist' who occasionally makes remarks which

can be misinterpreted as 'rationalist'. I am arguing that, on the contrary, Augustine is a thoroughgoing 'rationalist' who occasionally makes remarks which can be misinterpreted as 'voluntarist'.

Of course, anyone who dissents, as I am doing, from such a long and honourable tradition of interpretation as the 'voluntarist' reading of Augustine has work to do. The onus is on him, first, to justify his dissent by exegesis, and second, to explain how the traditional explanation arose.

I hope I have by now fulfilled the first part of this onus in this chapter and the last. In closing, I add some remarks in fulfilment of the second part of the onus.

8. From Augustine's Rationalism to Augustinian Voluntarism

Tracing the sources of what I take to be the mistake of classing Augustine, with (e.g.) Ockham, Scotus and Hume, as a thoroughgoing 'voluntarist', is no easy matter. To do it properly would call for a huge historical study which I have no intention of attempting here. But I think one may, at least, legitimately point to three errors of exegesis, and one of inference.

The error of inference is the mistake of supposing that, if (on Augustine's theory) renegation from any good action to wilful wrongdoing by 'mala voluntas' is possible, then, when any good

action is not reneged on but actually performed, a prior affirmation of that good action by 'bona voluntas' must be necessary.

The underlying image here (the 'intuition pump', as Dennett (1984) would say) is a common one in the philosophy of action (thanks to Plato's *Republic*). It is of the execution of an action as a *legislative process*. It is (we have seen) demonstrable that, in the case of a bad action, any process of means-ends reasoning which could lead to that action is neither necessary nor sufficient for its performance. What is both necessary and sufficient, for this bad intention to be promulgated as a bad action, is a kind of 'royal assent', a 'numen' or monarch's nod. And this, of course, is the contribution of the 'will'- which in the case of a bad action is so to speak 'ruling by decree'.

The inference is that if this 'royal assent' is necessary and sufficient in the case of a bad action, then it must be at least necessary in the case of *any* voluntary action at all. But why suppose this? To maintain the legislative analogy, there are many forms of constitution, and the royal assent is by no means an indispensable part of all of them. Why can't the correct analogy for good actions be a (republican) democratic constitution, where what is necessary and sufficient for the enactment of a law is simply that the constitutive assembly should (analogously to practical reasoning) debate and pass the resolution? In such a constitution, the royal assent is clearly neither necessary nor sufficient; it is irrelevant. And (still sticking with Plato!)

just because bad action is precisely that case where some tyrant desire rules by decree, why should we say that any τύραννος at all *must* be given a legislative role in any *ideal* constitution?

This idea, that 'voluntas' is involved as an extra component in every good or bad action, is suggested by the first error of exegesis, a misunderstanding of dCD 14.6:

'Voluntas est quippe in omnibus [his motibus (sc. the emotions)]; immo omnes nihil aliud quam voluntates sunt. Nam quid est cupiditas et laetitia nisi voluntas in eorum consensione quae volumus? Et quid est metus aut tristitia nisi voluntas in dissensione ab his quae nolumus?'

Note, in the first place, that Augustine is not here analysing *actions*, but *emotions*. More precisely, he is building a conceptual bridge between *emotions* and *attitudes*. 'Cupiditas', as an emotion, is a general wanting; as an attitude, it is a policy to obtain what is desired by the emotion. The difference between the emotion and the attitude is that attitudes are chosen, deliberately adopted, and voluntary, whereas emotions are not. So I am not (directly) morally responsible for my emotions; but I am directly responsible for my attitudes.

It is, in fact, a piece of over-enthusiasm which leads Augustine to say that 'Omnes [hi motus] nihil aliud quam voluntates sunt'. For he has just distinguished 'motus' ('emotion') rather carefully from 'voluntas' ('will'; or 'emotion

+ will' (= 'attitude')):

'Interest autem qualis sit voluntas hominis; quia si perversa est, perversa habebit hos motus.' (dCD 14.6)

The point is that one *can*, to a degree, be detached enough from one's emotions to choose whether or not to adopt their content into one's attitudes. That is why we are responsible for our attitudes, in a way in which we are not for our emotions. In any case the point, which is first made and then obscured by a rhetorical flourish on which far too much has traditionally been built, seems to have very little to do with the idea of the need for some separately identifiable assent of the 'voluntas' to any *action*.

The second error of exegesis concerns a passage in dCD 12.9 which has been misread:

'Cum ergo malae voluntatis efficiens naturalis nulla sit causa... Si dixerimus nullam esse efficientem causam etiam voluntatis bonae...'

The voluntarist interpretation of Augustine seems to me to require that this 'si' should be introducing the kind of rhetorical 'if' clause which is equivalent to an affirmative statement. The passage would then have Augustine saying that 'bona voluntas' *is*, just as much as 'mala', uncreated and without causal antecedents of any kind. But this is certainly not the

meaning of the passage, as simply reading on quickly shows:

'...Si dixerimus nullam esse efficientem causam etiam voluntatis bonae, cavendum est, ne voluntas bona angelorum non facta, sed Deo coaeterna esse credatur. Cum ergo ipsi facti sint, quo modo illa non esse facta dicetur?' (dCD 12.9)

The passage, in fact, says the exact opposite. The 'bona voluntas' of the good angels, unlike the 'mala voluntas' of the bad angels, *is* created, and *has* a causal antecedent, in that creative activity of God.

This enables us to deal with the third error of exegesis, which is the building of doctrines on a famous phrase which has already been quoted:

'Sed quae tandem esse poterit ante voluntatem causa voluntatis?' (dLA 3.49)

Taken on its own, this might well seem to be the claim that 'voluntas' *in general* has no antecedents of any sort, and would not be 'voluntas' if it had. But (as I have shown in Ch.5) to take it this way is to ignore the context of the remark, which makes it plain that it is only '*Improba voluntas*' which is under discussion here. Hence no general theory of what the antecedents of any 'voluntas' must be is here being established. Augustine's 'Sed quae tandem esse poterit ante voluntatem causa voluntatis?' should properly be read as 'Sed quae tandem esse poterit ante

malam voluntatem causa *malae* voluntatis?'. For, as we have already seen, Augustine explicitly says that 'bona voluntas' does have causal antecedents: in the creative activity of God. Compare dCD 12.9, above, with 5.9:

'Ac per hoc colligitur non esse causas efficientes omnium quae fiunt nisi voluntarias... In [Dei] voluntate summa potestas est, quae creatorum spirituum *bonas voluntates adiuvat, malas iudicat, omnes ordinat... Malae quippe voluntates ab illo non sunt, quoniam contra naturam sunt, quae ab illo est.*'

The implication is plainly that 'bonae voluntates', which (by definition) are not 'contra naturam', are from God.

That 'bona voluntas' has causal antecedents in the creative activity of God is also exactly what we should expect if I was right (Ch.8) about the Truth being the formal cause of the good will. And a third reason for taking this view is given by the clear asymmetry, in Augustine's works on grace, between attributions of praise (which are always to God) and attributions of blame (which are always to creatures). Again and again the refrain is: 'What have you that ye did not receive?' (v., e.g., dSL 58). This, again, is just what we should expect if 'bona voluntas', so far from being uncaused, comes directly from God.

My conclusion, then, is that Augustine is right to argue against the Manichaeans that, in his terms at least, 'bona voluntas' is radically asymmetrical to 'mala voluntas'. For 'bona

'voluntas' is explicable; it conforms to the form of the good life delineated in Ch.8; its aim is always some good or other; and it is even subject to the efficient causality of the Creator. 'Mala voluntas', on the other hand, is a surd in the equation, 'an incomprehensible given which steadfastly and in principle resists causal explanation' (Brown 1978, p.315); and it is deliberately left as such by Augustine (in his best moments). It partakes not of formal, nor of final, nor even of efficient causation; it is something which we simply cannot account for.

Moreover, it is to be noted that there is- in spite of the literature- no valid inference from the phenomenon of 'mala voluntas' as the tyranny of bad desire to a general voluntaristic hypothesis. It is a mistake to claim that there is or could be, in the anatomy of a 'normal' Augustinian voluntary action, some extra job of assenting which always needs to be done by the 'voluntas', over and above the simple choice and pursuit of a good. It is not the rational act which is the exception to the rule (however rarely that rule may be observed in a fallen world). The exceptional case is the irrational, inexplicable act of renegation which Augustine calls 'mala voluntas', and which gets the fallen world going in the first place.

9. Conclusion

What, then, has been established by my examination of Augustine's theory of freedom and the voluntary, and by

comparison of that theory with Aristotle's?

In the first place, I may hope to have justified my adoption, in my study of Augustine's theory, of the methodological axiom followed in my study of Aristotle: the axiom, namely, that showing the existence of voluntary action is sufficient and necessary for showing the existence of free action. If I am right about what is involved in voluntary action, on Augustine's best account (and in particular if voluntary action involves the ability to do otherwise than one does), this axiom has indeed been a useful and clarificatory one to follow, not only in the case of Aristotle but also in the case of Augustine. Augustine's theory of freedom is widely (and rightly) seen as a can of worms. If the methodological axiom which I have used is put to work, we can, in fact, extract a good deal of coherent and interesting theory from the plethora of Augustine's writings on the topic of freedom.

And secondly: it has emerged that there is a parallel problem in Augustine's theory to the problem posed by the apparent existence of *akrasia* for Aristotle's theory; a parallel problem, with a not quite parallel solution. Aristotle (I suggested in Ch.4) could have argued for the logical impossibility of full *akrasia*, meaning voluntary action which does not admit of rational explanation in the very strict terms of Aristotle's theory of practical rationality. Augustine has a looser theory of practical *wisdom* (rather than Aristotle's practical *rationality*), and a commitment to the teaching of Scripture that his analogue

of full akrasia, wilful wrongdoing, is nothing less than fully voluntary action. Hence it is that Augustine is bound to say that wilful wrongdoing- however mysterious, however inexplicable and however contrary in its direction to the directedness of practical wisdom towards the good- is, nonetheless, voluntary action, and therefore culpable even if not explicable. And so it turns out that, while voluntary action always and necessarily has three conditions for Aristotle- concerning ignorance, compulsion, and irrationality- for Augustine there is a counter example to these three postulated conditions. In the case of wilful wrongdoing we have voluntary action which is nonetheless irrational. This is the great difference between Augustine and his Classical predecessors. It is a difference which (I have argued) has been widely misinterpreted by the alignment of Aristotle and other classical writers with 'rationalist' understandings of the origination of action, and the alignment of Augustine with 'voluntarist' understandings thereof; as if Aristotle had been a proto-Kantian and Augustine a proto-Humean. But this difference between Aristotle and Augustine is, for all that it has been misrepresented, a profound and important one; and if the schematism under which I have attempted to analyse their theories of freedom and the voluntary has brought out that difference in a new and interesting light, that at least is a success of a kind.

Bibliography

Only books actually cited or referred to are included.

(i) Works of Aristotle, with abbreviations

- APo *Analytica Posteriora*
APr *Analytica Priora*
dA *de Anima*
dAR *de Arte Rhetorica*
dC *de Caelo*
dGA *de Generatione Animalium*
dI *de Interpretatione*
dMA *de Motu Animalium*
dPA *de Partibus Animalium*
EE *Ethica Eudemia*
MM *Magna Moralia*
Mph *Metaphysics*
NE *Ethica Nichomachea*
Phys *Physics*
Pol *Politics*
Top *Topics*
VV *Virtues and Vices*

(ii) Works of Augustine, with abbreviations

(The dates A.D. as given are not polemically asserted, and may well be only approximately right.)

- cAcad *contra Academicos* (386)
cDEP *cōtra Duas Epistolas Pelagianorum* (422-423)
cIOP *contra Iulianum Opus Imperfectum* (423)
Conf *Confessiones* (399-401)
dBV *de Beata Vita* (386)
dCD *de Civitate Dei* (413-426)
dCE *de Consensu Evangelistarum* (402)
dGCPO *de Gratia Christi et de Peccato Originali* (418)
dDA *de Duabus Animabus* (392-393)
dFRV *de Fide Rerum quae non Videntur* (400)
dGLA *de Gratia et Libero Arbitrio* (426)
dLA *de Libero Arbitrio* (387-388)
dMag *de Magistro* (389)
dME *de Moribus Ecclesiae* (388)
dNB *de Natura Boni* (404)
dPMR *de Peccatorum Meritis et Remissione* (411)
dPIH *de Perfectione Iustitiae Hominis* (415-416)
83DQ *de Diversis Quaestionibus LXXXIII* (388-395)
dQA *de Quantitate Animae* (387-388)
dSL *de Spiritu et Littera* (412)
dVR *de Vera Religione* (391)
dTrin *de Trinitate* (399-419)
EEGa1 *Expositio Epistolae ad Galatos* (394-395)

Ench	<i>Enchiridion</i> (423)
EPRoms	<i>Expositio LXXXIV Propositionum Epistolae ad Romanos</i> (394-395)
Retr	<i>Retractationes</i> (426-427)
RomsIE	<i>Epistolae Pauli ad Romanos Incohata Expositio</i> (394-395)
Sermon	<i>Sermons</i> (390-430)
Simp	<i>ad Simplicianum de Diversis Quaestionibus</i> (395)
Sol	<i>Soliloquia</i> (386-387)

Normally I have followed the Loeb text for works by Aristotle, and the text given in the series *Oeuvres de Saint Augustin* (La Bibliothèque Augustinienne) for works by Augustine. I have also consulted the Oxford Classical Text of Aristotle (*Nicomachean Ethics* edited by Ingram Bywater, 1890), and the Loeb edition of Augustine's *Confessions*.

(iii) Other works

Ackrill (1965): J.L.Ackrill, 'Aristotle's Distinction between *Energeia* and *Kinesis*': p.121-142 in Bambrough (1965).

Anscombe (1957): G.E.M.Anscombe, *Intention*. Oxford: Clarendon Press, 1957.

Anscombe (1965): G.E.M.Anscombe, 'Thought and Action in Aristotle: What is "Practical Truth"?' In Bambrough (1965).

Austin (1956/7): J.L.Austin, 'A Plea for Excuses', *Proceedings of the Aristotelian Society*, 57 (1956/7), p.1-30.

Austin (1961): J.L.Austin, 'Ifs and Cans', in J.O.Urmson and

- G. Warnock, eds., *The Philosophical Papers of J.L. Austin* (Oxford: Clarendon Press, 1961), p.153-180.
- Bambrough (1965): Renford Bambrough, ed., *New Essays on Plato and Aristotle*. London: Routledge Kegan Paul, 1965.
- Bible, Authorised Version of, 1611.
- William Blake, *The Marriage of Heaven and Hell*. In J. Sampson, ed., *The Poetical Works of William Blake*. Oxford: Oxford University Press, 1913.
- Boethius, Anicius Manlius Severinus, *Consolatio Philosophiae*. Text, and translation (1609) by 'I.T'., in Loeb Classical Library (London: Harvard University Press, 1936).
- Brown (1978): Robert Brown, 'The First Evil Will must be Incomprehensible: a Critique of Augustine', *Journal of the American Academy of Religion*, 46 (1978), p.315-329.
- Burleigh (1953): J.H.S. Burleigh, *Augustine: Earlier Writings, Selected and Translated*. London: S.C.M. Press (Library of Christian Classics), 1953.
- Burnaby (1938): John Burnaby, *Amor Dei: a Study of the Religion of St. Augustine*. London: Hodder and Stoughton, 1938.
- Burnet (1900): John Burnet, *The Ethics of Aristotle, Edited with an Introduction and Notes by John Burnet*. London: Methuen, 1900.
- Burnyeat (1987): Myles Burnyeat, 'Wittgenstein and Augustine *de Magistro*', *Proceedings of the Aristotelian Society* (Supplementary Volume), 1987.
- Charles (1984): David Charles, *Aristotle's Theory of Action*. Ithaca: Cornell University Press, 1984.
- Charlton (1988): William Charlton, *Weakness of the Will: a Philosophical Introduction*. Oxford: Clarendon Press, 1988.

- Cicero, *Tusculanae Disputationes*. Text and translation by J.E.King in Loeb Classical Library (London: Harvard University Press, 1927).
- Clark (1978): S.R.L.Clark, *Aristotle's Man*. Oxford: Clarendon, 1978.
- Davidson (1980): Donald Davidson, *Essays on Actions and Events*. Oxford: Clarendon, 1980.
- Dennett (1984): Daniel Dennett, *Elbow Room: the Varieties of Free Will Worth Wanting*. Oxford: Clarendon, 1984.
- René Descartes, *Oeuvres de Descartes*, edd. Adam and Tannery. Paris: Leopold Cerf, 1897-1913.
- DK: Diels and Kranz, *Fragmente der Vorsokratiker*, 3 vols. Berlin: Weidmannsche Verlagsbuchhandlung, 1937.
- Dihle (1982): Albrecht Dihle, *The Theory of Will in Classical Antiquity*. Berkeley: University of California Press, 1982.
- Foot (1981): Philippa Foot, *Virtues and Vices and other Essays in Moral Philosophy*. Oxford: Clarendon Press, 1981.
- Gardiner (1969/70): Patrick Gardiner, 'Error, Faith and Self-Deception', *Proceedings of the Aristotelian Society*, NS 70 (1969/70), p.221-244.
- Gauthier-Jolif (1970): Gauthier and Jolif, *L'Ethique à Nichomaque*. Louvain: Publications Universitaires, 1970.
- Geach (1977): Peter Geach, *The Virtues*. Cambridge: Cambridge University Press, 1977.
- Gilson (1960): Etienne Gilson, *The Christian Philosophy of Saint Augustine*. London: Victor Gollancz, 1960.
- Gosling (1973/4): J.C.B.Gosling, 'More Aristotelian Pleasures', *Proceedings of the Aristotelian Society*, NS 74 (1973/4), p.15-34.

- Gosling (1990): J.C.B.Gosling, *Weakness of the Will*. London: Routledge, 1990.
- Gosselin (1949): Gosselin, annotated French edition and translation of dNB (in *Oeuvres de Saint Augustin* series, vol.1). Paris: Desclée, de Brouwer et Cie, 1949.
- Hardie (1980): W.F.R.Hardie, *Aristotle's Ethical Theory*, Second Edition. Oxford: Clarendon Press, 1980.
- Hare (1952): Richard Hare, *The Language of Morals*. Oxford: Clarendon Press, 1952.
- Hare (1963): Richard Hare, *Freedom and Reason*. Oxford: Clarendon Press, 1963.
- Hare (1971): Richard Hare, 'Practical Inferences', in Richard Hare, *Practical Inferences*. London: Macmillan, 1971.
- Hume (1739): David Hume, *A Treatise of Human Nature* (1739). Ed. L.A.Selby-Bigge, Oxford: Clarendon, 1964.
- Hume (1775): David Hume, *An Enquiry Concerning Human Understanding* (1775). Ed. L.A.Selby-Bigge, Oxford: Clarendon, 1975.
- Irwin (1987): Terence Irwin, *Aristotle's First Principles*. Oxford: Clarendon Press, 1987.
- Joachim (1955): H.H.Joachim, *Commentary on the Nichomachean Ethics*, ed. D.A.Rees. Oxford: Clarendon Press, 1955.
- Justin Martyr, *Dialogus contra Tryphonium*.
- Immanuel Kant, *Grundlegung zur Metaphysik der Sitten*, translated by H.J.Paton as *The Moral Law* (London: Hutchinson University Library, 1971).
- Keillor (1989): Garrison Keillor, 'The Current Crisis in Remorse', in *We Are Still Married* (London: Faber, 1989).

- Kenny (1979): Anthony Kenny, *Aristotle's Theory of the Will*. London: Duckworth, 1979.
- Kirwan (1989): Christopher Kirwan, *Augustine*. London: Routledge (*The Arguments of the Philosophers Series*), 1989.
- Lemmon (1962): E.J.Lemmon, 'Moral Dilemmas', *Philosophical Review*, 71 (1962), p.139-158.
- Long (1974): A.A.Long, *Hellenistic Philosophy*. London: Duckworth, 1974.
- MacIntyre (1985): Alasdair MacIntyre, *Whose Justice? Which Rationality?*. London: Duckworth, 1985.
- MacIntyre (1990): Alasdair MacIntyre, *Three Rival Versions of Moral Inquiry*. London: Duckworth, 1990.
- Mackie (1982): J.L.Mackie, *The Miracle of Theism*. Clarendon: Oxford, 1982.
- Molina (1588): Luis de Molina, *Liberi Arbitrii cum Gratiae Donis, Divina Praescientia, Providentia, Praedestinatione et Reprobatione Concordia*, ed. J.Rabeneck. Madrid, 1953.
- Moravscik (1967): Julius Moravscik, ed., *Aristotle: a Collection of Critical Essays*. Garden City, 1967.
- Murdoch (1970): Iris Murdoch, *The Sovereignty of Good*. London: Routledge Kegan Paul, 1970.
- Novum Testamentum Graece*, edd. Nestlé/ Aland. Stuttgart: Teubner, 1898/ 1979.
- Friedrich Nietzsche, *Götterdämmerung*. Translated by R.J.Hollingdale as *The Twilight of the Idols* (London: Harmondsworth, 1968).
- Nussbaum (1986): Martha Craven Nussbaum, *The Fragility of Goodness*. Cambridge: Cambridge University Press, 1986.

Owen (1967): G.E.L.Owen, 'τιθέσθαι τὰ φαινόμενα', in Moravcsik (1967).

Pike (1965): Nelson Pike, 'Divine Omniscience and Human Freedom', *Philosophical Review*, 74 (1965), pp.27-46.

Plantinga (1976): 'On Ockham's Way Out', reprinted in T.V.Morris, ed., *The Concept of God*. Oxford: Oxford University Press (Oxford Readings in Philosophy), 1987.

Plato, *Laches, Protagoras, Meno, Euthydemus*. Text and translation by W.R.M.Lamb in Loeb Classical Library (London: Harvard University Press, 1924).

Plato, *Republic*. Text and translation (2 vols.) by Paul Shorey in Loeb Classical Library (London: Harvard University Press, 1935).

Plotinus, *Enneads*. Text and translation by A.H.Armstrong in Loeb Classical Library. 8 vols. London: Harvard University Press, 1978).

Rackham (1926): *Aristotle's Nichomachean Ethics*. Text and translation by H.Rackham in Loeb Classical Library (London: Harvard University Press, 1926).

Raz (1978): Joseph Raz, ed., *Practical Reasoning*. Oxford: Clarendon (Oxford Readings in Philosophy), 1978.

Ricoeur (1967): Paul Ricoeur, *The Symbolism of Evil*. New York: Harper and Row, 1967.

Rist (1969): John Rist, 'Augustine on Free Will and Predestination', *Journal of Theological Studies*, N.S. 20 (1969), p.420-447.

Rorty (1980a): Amélie Oksenberg Rorty, 'Akrasia and Pleasure: Nichomachean Ethics Book VII', in Rorty (1980b), p.267-283.

Rorty (1980b): Amélie Oksenberg Rorty, ed., *Essays on Aristotle's Ethics*. London: University of California Press, 1980.

Ross (1925): W.D.Ross, *Oxford Translation of Aristotle*. Oxford: Oxford University Press, 1925.

Ryle (1949): Gilbert Ryle, *The Concept of Mind*. London: Hutchinson University Library, 1949.

Sartre (1945): J.-P.Sartre, *Huis Clos: Pièce en Un Acte*. London, 1945.

Sartre (1949): J.-P.Sartre, *L'Être et le Néant: Essai d'Ontologie Phénoménologique*. Paris, 1949. (Translated by Hazel Barnes as *Being and Nothingness*.)

Stoutland (1970): 'The Logical Connection Argument', *American Philosophical Quarterly*, 1970.

Stoutland (1976): 'The Causation of Behaviour', in *Essays in Honour of G.H. von Wright*, Amsterdam 1976.

Strawson (1962): 'Freedom and Resentment', *Proceedings of the British Academy*, 1962.

Thonnard (1941): F.J.Thonnard, annotated French edition and translation of dLA (in *Oeuvres de Saint Augustin* series, vol.6). Paris: Desclée, de Brouwer et Cie, 1941.

Tolstoy, *War and Peace* (1867). Translated Rosemary Edmonds (Penguin Classics; London: Harmondsworth, 1961).

Wallis (1887): Robert Wallis, translation of cDEP, in *Augustine's Anti-Pelagian Writings*; vol.5 of Schaff, ed., *The Select Library of the Nicene and Post-Nicene Fathers* (Grand Rapids, Michigan: Eerdmans, 1887/ 1971), p.374-435.

Wiggins (1980): David Wiggins, 'Deliberation and Practical Reason', in Rorty (1980b), p.221-240.

Wittgenstein (1967): Ludwig Wittgenstein, *Philosophical Investigations*, edd. G.E.M.Anscombe and Rush Rhees, with an English translation by G.E.M.Anscombe. Oxford: Basil Blackwell, 1967 (third edition).

von Wright (1978): Georg Henrik von Wright, 'On So Called Practical Inference'. In Raz (1978), p.46-62.