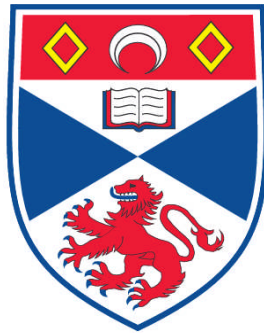# ARISTOTLE ON ETHICAL ASCRIPTION:
A PHILOSOPHICAL EXERCISE IN THE INTERPRETATION OF THE ROLE AND SIGNIFICANCE OF THE *HEKOUSIOS/AKOUSIOS* DISTINCTION IN ARISTOTLE'S ETHICS

Javier Echeñique

**A Thesis Submitted for the Degree of PhD
at the
University of St. Andrews**

2010

ARISTOTLE ON ETHICAL ASCRIPTION:

A philosophical exercise in the interpretation of the role and significance

of the *hekousios* / *akousios* distinction in Aristotle's Ethics.

Javier Echeñique

Thesis for the degree of

Doctor of Philosophy (Philosophy)

University of St Andrews

April 2010

Supervisors:

Professor Sarah Broadie (Philosophy)

Dr. Jens Timmermann (Philosophy)

1. Candidate's declarations:

I, Javier Echenique, hereby certify that this thesis, which is approximately 84.000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in 09/2006 and as a candidate for the degree of Ph.D (Philosophy) in 09/2006; the higher study for which this is a record was carried out in the University of St Andrews between 2006 and 2010.

Date: 06/07/2010   signature of candidate :

2. Supervisor's declaration:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of ... ...PhD.................... in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date: 8 July, 2010   signature of supervisor:

3. Permission for electronic publication: (*to be signed by both candidate and supervisor*)

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Access to printed copy and electronic publication of thesis through the University of St Andrews.

Date:08/07/2010   Signature of supervisor:

                    Signature of candidate

ABSTRACT

In his ethical treatises Aristotle offers a rich account of those conditions that render people's behaviour involuntary, and defines voluntariness on the basis of the absence of these conditions. This dissertation has two aims. One is to offer an account of the significance of the notions of involuntariness and voluntariness for Aristotle's ethical project that satisfactorily explains why he deems it necessary to discuss these notions in his Ethics. My own account of the significance of these notions for Aristotle's Ethics emerges from my arguments against the two most influential views concerning this significance: I argue that Aristotle's concern with voluntariness in his Ethics is not (primarily) shaped by a concern with accountability, i.e. with those conditions under which fully mature and healthy rational agents are *held accountable* or *answerable* for their actions; nor is it (primarily) shaped by a concern with the *conditioning* of pain-responsive agents for the sake of socially useful ends that are not, intrinsically, their own. Rather, his concern is with reason-responsive agents (which are not morally accountable agents, nor merely pain-responsive agents) and the conditions for attributing ethically significant behaviour to them. This is what I call 'ethical ascription'. The second aim of this dissertation is to provide a comprehensive account of those conditions that defeat the ascription of ethically significant pieces of behaviour to reason-responsive agents, and to show the distinctiveness of Aristotle's views on the nature of these conditions. The conclusions I arrive at in this respect are shaped by the notion of ethical ascription that I develop as a way of reaching the first aim.

ACKNOWLEDGEMENTS

Abbreviations of Aristotle's texts used in this dissertation

Analytica Posteriora: *An. Post.*

Analytica Priora: *An. Pr.*

Ars Rhetorica: *Rhet.*

Categoriae: *Cat.*

De Anima: *De An*.

De Caelo: *Cael*.

De Generatione et Corruptione: *GC*

De Sophisticis Elenchis: *Soph. El.*

De Virtutibus et Vitiis: *De Virt*. (spurious)

Ethica Eudemia: *EE*

Ethica Nicomachea: *EN*

Magna Moralia: *MM* (disputed authorship)

Physica: *Phys*.

Poetica: *Poet*.

Politica: *Pol*.

Topica: *Top*.

TABLE OF CONTENTS

Any derogatory criticism, say of the workmanship of a product or the design of a machine, can be called blame or reproach. So we want to put in the word "morally" again: sometimes a failure may be *morally* blameworthy, sometimes not. Now has Aristotle got this idea of *moral* blame, as opposed to any other? If he has, why isn't it more central?

<div align="right">

G. E. M. Anscombe

'Modern moral philosophy'

</div>

INTRODUCTION

A. Reputable opinions

In his 1960 book, *Merit and Responsibility*, in all probability the first systematic effort in the English language to closely scrutinize the development of the notion of moral responsibility in Ancient Greek culture, Arthur W. H. Adkins, in the course of analyzing Aristotle's own views on this elusive topic, famously concluded that Aristotle fails to develop a satisfactory concept of moral responsibility, for the same reasons as Homer did. Two premises ground this severe and influential verdict, the first being the main thesis of Adkins' book, namely, that the "chief impediment in the way of the development of a satisfactory concept of moral responsibility in Greek" (p. 332) was the traditional concept of *arête* (virtue) a concept that "was inimical to the concept of moral responsibility" (p. 337). The second premise is simply the application of this general thesis to Aristotle: Aristotle inherits this traditional concept of *aretê*; this concept embraces, in addition to what Adkins terms the 'quiet', 'co-operative' virtue of justice (the 'new' excellence), those excellences traditionally valued by the Greeks from Homeric times, like courage, liberality, etc.

Why exactly did Adkins think that the traditional concept of *aretê* was inimical to the concept of moral responsibility? If I interpret him correctly, he thought this because he believed that the traditional concept of *aretê* was un-Kantian, whereas the concept of moral responsibility is distinctively Kantian: "For any man brought up in a western democratic society the related concepts of duty and responsibility are the central concepts of ethics; and we are inclined to take it as an unquestionable truth, though there is abundant evidence to the contrary, that the same must be true of all societies. In this respect at least we are all Kantians now" (p. 2). That the Greeks were not Kantians is an uninteresting claim. That the concept of moral responsibility is distinctively Kantian, on the other hand, because closely bound to the concept of duty or obligation, begins to sound more interesting. As a matter of fact, it contains the truth that I wish to preserve of Adkins' chief thesis.

This truth, I am now convinced, could only have been fully appreciated after five decades of intense philosophical discussion on the subject of moral responsibility. Such intense discussion, initially stimulated by P. F. Strawson's 1962 article 'Freedom and Resentment' (published just two years after *Merit and Responsibility*), has provided us with a series of insights, some of which are particularly relevant to Adkins' thesis. One of these insights is (1) that the concept of duty or obligation (Strawson's 'moral expectation') is indeed closely connected to the concept of moral responsibility *as accountability* or *answerability* - and we are nowadays in a much better position than Adkins was, to understand the ways in which these two notions of *obligation* and *accountability*, along with a whole cluster of concepts like *desert*, *sanction*, etc., are connected with each other. I will argue in Chapter 1 that, if Adkins' chief claim amounts to the thesis that Aristotle is not particularly concerned with a conception of moral responsibility as accountability, he is right. Whether this is correctly characterized as a 'failure' on Aristotle's part, however, is a different matter, and one about which I will have something to say further on.

Now, that an agent S is morally responsible for a piece of behaviour *B* if *B* is relevant to S's moral assessment or evaluation, and that moral responsibility for pieces of behaviour (or their consequences) is in this way closely tied to the moral assessment of their agents, seems to me a tolerably uncontroversial truth about the nature of moral responsibility (although a pure consequentialist may deny it). Moreover, it seems to me that philosophers have always acknowledged, at least tacitly, that a given conception of moral responsibility will *depend* on what we take the nature of this assessment or evaluation to be, and that the nature of this assessment will in turn determine the criteria (including what sort of agent is a fitting object of moral assessment) on the basis of which a given piece of behaviour is deemed relevant to the assessment of an agent. The insight peculiar to the post-strawsonian debate in this regard is (2) the increasing recognition that - what we may call - the 'instrumentalist account' of our praising and blaming practices cannot do justice to the moral nature of those assessments involved in genuinely moral evaluation, and accordingly that this account cannot ground a conception of moral responsibility. Before Strawson's paper, some philosophers like J. J. C. Smart argued

that in a deterministic system our *moral* practices of praise and blame could still be salvaged, because the primary role of these practices is to grade the agent with a view to letting *others* know what the agent is like, in ways that can be *socially useful*.[1] After Strawson, philosophers widely agree that this socially useful, purely instrumentalist grading, fails to do justice to the moral nature of our praising and blaming practices. The Strawsonian argument underpinning this agreement is that this purely socially useful grading 'instrumentalizes' the sane, adult human agent and fails to acknowledge him[2] as a moral agent in his own right (i.e. the 'instrumentality argument', as I will call it).

To this instrumentalist account of our praising and blaming practices, and the sort of evaluation involved in them, Strawson and those philosophers influenced by his approach, oppose what may be called the 'Strawsonian account' of moral praise and blame. According to Strawsonians, genuinely moral praise and blame bestowed on a sane human agent for a given piece of behaviour B (or its consequences), essentially involve reactive attitudes (or at least the belief that these are appropriate) like moral indignation, resentment, respect or admiration, held towards the agent on the occasion of his voluntarily causing B. Essential to this conception of moral responsibility is (a) the view that the operation of the reactive attitudes requires that their appropriate target be a morally accountable agent (a notion I will define in Chapter 1), and also (b) the conviction that the 'reactive' and the instrumentalist accounts of praise and blame are disjunctive alternatives, that is, that praise and blame are either reactive attitudes or instrumentalist (i.e. 'objective' in Strawson's terminology).

The valuable insight of the post-strawsonian debate, according to which purely socially useful grading fails to do justice to the moral nature of our moral praising and blaming practices, is, I think, obscured by the strong tendency, already detected in Strawson's article, to classify *educational* praising and blaming practices under the rubric of 'instrumentalist' or 'objective' attitudes. Central to my argument

---

[1] See Smart, J. J. C. (1961). On Smart's view, the agent is typically graded on the occasion of his *voluntary* actions, primarily because only these actions tell others what the agent is like, but also because he thinks praise and blame have a secondary function, which is to encourage the agent to, or discourage him from, repeating the action, and this requires some level of control over the action.

[2] When referring to the agent in my examples, or Mr. S, I will be using the masculine pronoun.

is a distinction between *instrumentalist* praise and blame, that is, the sort of praise and blame that is ruled out as non-moral on the basis of the instrumentality argument, and the wider category of *prospective* or forward-looking praise and blame, which includes instrumentalist (or 'objective') attitudes, *as well as* the praising and blaming involved in moral education, which – I argue – is not subject to the instrumentality argument. The Strawsonian mistake is to bring all prospective evaluative attitudes under the rubric of 'instrumentalist' (or Strawson's 'objective') and therefore 'non-moral' (due to the instrumentality argument). As a result of this fusion, the Strawsonian wrongly categorizes all targets of prospective attitudes (animals, psychopaths *as well as* children) under the class of non-moral agents, and wrongly deems all non-prospective attitudes (which can now be called 'retrospective') as exclusively 'moral'.

The preceding remarks on what I think are the shortcomings of Strawson's view are crucial to my present purposes, because my argument in Chapter 1 and 2 depends on showing that the two most influential post-strawsonian interpretations of Aristotle's theory of voluntariness in his Ethics inherit the flaws of the Strawsonian account; in particular, the fault of dismissing educational attitudes as non-moral, and thus pupils in the process of moral education as non-moral agents. As it is natural, these two (opposed) post-strawsonian views are ultimately based on an interpretation of the sort of praise and blame which, as Aristotle says, is bestowed only upon voluntary responses (*EE* 1223a9-15; *EN* 1109b30-34). One such interpretation is what I call the 'prospective conditioning interpretation' about the significance of the involuntary/voluntary distinction for Aristotle's Ethics.[3] According to this interpretation, the sort of praise and blame Aristotle says is appropriate to voluntary responses is *prospective* in nature.[4] One of the mistakes of the prospective conditioning interpretation, inherited from the Strawsonian view, is to conflate 'prospective' with 'instrumentalist', and therefore to dismiss *all* prospective attitudes as irrelevant for the purposes of moral evaluation, along with all the recipients of prospective attitudes (animals, psychopaths *and* children alike) as non-

---

[3] From now on I will use the term 'Ethics' or 'Aristotle's Ethics' to refer both to the *EN* and the *EE*. I will occasionally make use of the *Magna Moralia* as well, but I hope none of my interpretations heavily depends on attributing this work to Aristotle himself.
[4] This view is chiefly represented by Roberts, J. (1989).

moral agents. Since (i) both animals and children, according to Aristotle, act voluntarily (*EN* 1111a25-6), then, on this view (ii) the sort of praise and blame Aristotle is interested in must be prospective, and (iii) since all prospective attitudes are instrumentalist (on the mistaken, Strawsonian view), then (iv) these are not moral attitudes, and their recipients are not moral agents. This, I think is the argument underlying Jean Roberts's claim that Aristotle's notion of voluntariness (as it occurs in his Ethics) "does not coincide ... with any later notion of moral responsibility". According to Roberts, "whatever precisely moral responsibility is, animals and small children do not have it".[5] So Roberts, just as Adkins, thinks that Aristotle does not have a theory of moral responsibility.

Yet another line of interpretation, what I call the 'Strawsonian interpretation', is firmly opposed to this conclusion. This view is maintained by Terence Irwin in his 1980 paper 'Reason and responsibility in Aristotle', and further elaborated by Susan Sauvé-Meyer's 1993 book *Aristotle on Moral Responsibility.* Both authors think that Aristotle *does* have a theory of moral responsibility, and furthermore, one that is, in some important respects, distinctively modern (i.e. Strawsonian). According to Sauvé-Meyer, the praise and blame Aristotle is interested in is not prospective, but *retrospective*: "Aristotle intends his account of voluntariness" – she says – "to capture the conditions in which an agent's states of character are praiseworthy and blameworthy. And these conditions ... are those in which an agent is causally responsible for particular good and bad actions or feelings. The sort of praise and blame the agent deserves for ... these actions and feelings is therefore based on a *retrospective evaluation – for the question is whether the agent's character produced them*."[6]

Sauvé-Meyer thinks that Aristotle is interested in the conditions of voluntariness in his Ethics, because he is interested in the causal conditions in which adults in full possession of virtue- or vice-dispositions (broadly understood) are non-accidentally productive of morally significant outcomes, and these causal conditions correspond to the conditions of voluntariness. Since neither children nor animals possess ethical dispositions, then it turns out that Aristotle is interested in

---

[5] *Ibid.*, p. 25.
[6] Sauvé-Meyer, S. (1994), p. 51. Italics are mine.

retrospective evaluations of mature ethical agents in possession of ethical dispositions. By 'retrospective' Sauvé-Meyer means that these are evaluations of ethically (or 'morally' in her use) significant pieces of behaviour that can be *traced back* to the corresponding ethical dispositions (and the agents in possession of them) as their *causal antecedents*. She finds no problem with identifying *this* concern with the conditions of voluntariness, with a concern with the conditions of *moral responsibility*.

The chief assumption of the conception of moral responsibility assumed by Sauvé-Meyer is that genuinely *moral* evaluations are retrospective or backward-looking, but why? She in particular emphasizes the *causal antecedent* of an ethically significant response in an agent's character: but how is this emphasis on the causal antecedent supposed to help us understand why our evaluating the mature agent for having caused that response is a moral evaluation, in a substantive sense of 'moral'? It is clear that Sauvé-Meyer and Irwin rely (sometimes implicitly, other times explicitly) on a Strawsonian conception of moral responsibility, and that this conception furnishes the ultimate justification for the view that genuinely moral evaluations are *retrospective*. This view, I am now convinced, is the result of the Strawsonian conflation between 'prospective' and 'instrumentalist' (or 'objective') on the one hand, and 'retrospective' and 'reactive' (and therefore exclusively 'moral') on the other, and the consequent exclusion of *all* prospective attitudes as irrelevant for the purposes of moral evaluation, along with their appropriate recipients (animals, psychopaths *and children* alike) as non-moral agents. As a consequence, the Strawsonian interpretation of Aristotle wrongly imports into Aristotle's Ethics the distinctively modern concern with the notion of the *moral agent* in the particular sense of *morally accountable* or *answerable agent*, i.e. the appropriate target of what Strawson called 'reactive attitudes'. It seems to me that Aristotle's proharetic agent,[7] the mature agent in possession of a full ethical disposition, *coincides* with this moral agent (that is, coincides as far as the *extension* of these concepts goes), but he is not this moral agent, and that if one desists from

---

[7] By 'prohairetic agent' I mean the fully mature, healthy human being in possession of a fully developed ethical disposition and capable of *prohairesis*, a reasoned choice based on ethical deliberation.

taking the Strawsonian picture of moral responsibility as a model for the interpretation of Aristotle, there is no reason to think that the prohairetic agent is the *exclusive* target of (Aristotelian) praising and blaming *logoi*, a praise and a blame which – I am also convinced – cannot be understood in terms of the Strawsonian reactive attitudes.

These ideas furnished the starting point for this dissertation. Somehow, however, it seemed to me that there was *some* truth both in Robert's claim that Aristotelian praise and blame were forward-looking, and in Sauvé-Meyer's insistence on the prohairetic agent being the main concern of Aristotle's Ethics, its main focus.[8] The truth in these two accounts, I take it, is preserved in my own interpretation of Aristotle's concern with the voluntary/involuntary distinction in his Ethics, developed in Chapter 2. This interpretation also explains why Adkins was, after all, wrong to think that Aristotle 'fails' to develop a concept of responsibility *qua* accountability. Let me explain.

Another insight of the post-strawsonian debate, (this time, Gary Watson's), turned out to be extremely significant: (3) accountability (i.e. an agent's being held morally responsible or accountable for complying with certain demands) is not *all* there is to our concept of moral responsibility, or better, our concept of 'moral praise and blame'. There are certain ethical assessments that we make about agents on the basis of their conduct that are made from a different perspective and do not (at least not *per se*) invoke the cluster of concepts associated with accountability. This perspective has been labelled by Gary Watson 'aretaic' (from '*aretê*', virtue), and it is the perspective from which we pass a moral judgement on some agent but are not thereby holding him accountable. Rather, by passing this judgement we are merely *attributing* a certain conduct to a moral fault in him; this is what Watson calls 'attributability' (i.e. the attribution of an action to an agent). It seemed to me that my criticisms to the Strawsonian interpretation (and, secondarily, to the prospective conditioning interpretation) in Chapter 1, and my analysis of Aristotle's own notions of praise in the same chapter, strongly suggested that Aristotle's concerns were better modelled on a conception of responsibility as attributability. I do this in

---

[8] Sauvé-Meyer, S. (1994), Chapter 2.

Chapter 2. Because of this third insight, Adkins was simply wrong to think that Aristotle failed to develop a concept of moral responsibility *tout court.*

I argue in Chapter 2 that Aristotle's conception of moral responsibility can be partly understood as attributability, but I also argue that it has some distinctive features of its own, for which reason I have chosen to label Aristotle's own version of attributability, 'ethical ascription', and I will talk of 'ascription' and 'ascribe' to mark the fact that these distinctive features are not fully covered by Watson's discussion of the aretaic perspective. Now, one of the chief features peculiar to ethical ascription is the fact that virtue- and vice-terms – and the corresponding praising and blaming *logoi* – have a focal use and a derivative, secondary use: the account of praising and blaming *logoi* based on the latter distinction captures, it seems to me, the truth contained both in the prospective conditioning and in the Strawsonian interpretation. I argue in Chapter 2 that, in their focal meaning or use, praise and blame *logoi* apply to the prohairetic agent: when I call someone S 'just' or 'cowardly' on the occasion of his recognizably just or cowardly behaviour, I am not necessarily implying that S is in full possession of the virtue of justice or the vice of cowardice, but rather invoking this full virtue, or the concept of an agent in full possession of it, by saying that S is *like* him. It is this focal meaning of praising and blaming *logoi* that accounts for the central role of the prohairetic agent in Aristotle's Ethics, so emphasized by the Strawsonian account. Moreover, this account of praising and blaming *logoi* as having both focal and secondary or derivative meaning, has the advantage of making room for children within the domain of application of praise and blame, and thus captures the truth contained in the prospective conditioning account.

This domain of moral agency, I further argue in Chapter 2, is best characterized by the notion of reason-responsiveness. The reason-responsive agent corresponds to the agent to whom ethically significant items can be ascribed; since children are amongst reason-responsive agents, the reason-responsive agent is not the morally accountable agent (the appropriate target of the S-reactive attitudes), nor is he the agent whose behaviour can be conditioned or manipulated by means of pain and pleasure, punishment or rewards (the pain-responsive agent, a category that includes non-rational animals and irreversible psychopaths).

B. The negative method

This dissertation has two parts, and the preceding section pretty much summarizes the contents of the first, which I call 'Towards an account of ethical ascription' (Chapters 1 and 2). The second part is concerned with the defeaters of ethical ascription (Chapters 3 to 7). In Chapter 2 I conclude that Aristotle is interested in the conditions of voluntariness in his Ethics, because he is interested in the following question:

Qe: What are the conditions that need to be fulfilled in order for an apt observer to be warranted in (1) praising and blaming an action A as 'v' and (2) praising or blaming a reason-responsive agent S in 'V'-terms through (*dia*) (1)?

As I will argue in Chapter 7, conditions (1) and (2) in Qe are two logically distinguishable steps in ethical ascription. An apt observer may be warranted in praising or blaming a given action A in virtue- or vice-terms or as grounding such description (i.e. describing it in 'v'-terms, for short), if A has been voluntarily performed, and still not be warranted in praising or blaming S as 'V' through his voluntarily doing of A. For the latter to be the case, a further condition needs to be met by S's moral attitudes towards A, that is not entailed by the fact that A has been voluntarily performed; namely, the lack of moral pain condition. I will call these two conditions in Qe, the conditions of ethical ascription or alternatively, the conditions of praise and blame: they correspond to the conditions of voluntariness (i.e. (1)), whatever these happen to be, plus the lack of what I will call, in Chapter 7, 'moral pain' (i.e. (2)).

The conditions of praise and blame are not to be confused with the narrower set of conditions that, according to Aristotle, an action A (e.g. refusing to eat an enormous cake) must meet in order for it to be the *manifestation* of an already developed, 'perfect' ethical disposition. According to this narrower set of conditions, A needs *not only* (i) be voluntarily performed (e.g. done by yourself in full knowledge that what you are refusing is to eat an enormous, a recognizably temperate action) and (ii) done without the relevant sort of pain, and perhaps with pleasure (*EN*

1104b5-8); A must *also* (iii) be the result of *prohairesis*, a considered judgement or decision to the effect that A (e.g. refusing to eat that enormous cake) is, overall, good/bad conduct (*EN* 1105a31), and (iv) good/bad conduct *qua* A (*EN* 1105a32) (e.g. refusing to eat it *qua* recognizably temperate in the circumstances, as opposed to *qua* disappointing your friends by so refusing).

Now, conditions (iii) and (iv) can only be fulfilled by fully mature, healthy rational agents in possession of a perfect ethical disposition.[9] Because of this, they are far too stringent as conditions of ethical ascription. (The reasons for this are, first, that the flexibility afforded by distinguishing a focal and a secondary meaning of praising and blaming *logoi* allows them to be applied to children; and second, that incontinent agents are also within the domain of ethical ascription, and incontinence is not a *perfect* (*teleia*) ethical disposition). Since children cannot act on *prohairesis* (*EN* 1111b9), nor can incontinent agents as such,[10] then ethical ascription does not require that the agent act on *prohairesis* in order for his act to be ascribed to him in 'v'-terms.

The conditions of ethical ascription involve (i) and (ii). But before commenting on these conditions, it is important to rule out another condition that may be thought to be relevant to the concept of moral responsibility, and which some scholars have regarded as central to Aristotle's account of voluntariness. Some scholars, most notably David J. Furley, have argued that Aristotle's argument for the voluntary acquisition of ethical dispositions in *EN* III 5 plays a central role in Aristotle's theory of responsibility because "[f]rom a moral point of view, an act is voluntary if it proceeds from a disposition which is voluntary."[11] Furley thinks that if an act is to be praised or blamed in 'v'-terms (and its agent praised or blamed through it), such an act must result from the exercise of a *voluntarily* acquired disposition, and that this is a requirement arising from the 'moral point of view'. It is not unlikely that the 'moral point of view' Furley adopts, and from which his stringent requirement stems, is accountability. Indeed, as we shall see in Chapter 2,

---

[9] From now on, 'full' or 'perfect' ethical dispositions are not meant to include continence and incontinence.

[10] Although they act voluntarily (*EN* 1152a15-19). Aristotle thinks that the incontinent is not fully faulty because his *prohairesis* is good; it is just that he fails to act on it.

[11] Furley, D. J. (1967), p. 191

there is some plausibility in the idea that paradigmatic Strawsonian reactive attitudes like resentment or moral indignation (and sanctioning dispositions in general) are withheld as soon as their target is seen as someone who has had no control whatsoever over the formation of his own (flawed) character and desires. This is not the point of view of ethical ascription, however. From the latter point of view, the *possession* of an ethical disposition is not a necessary condition of ascribing an ethically significant response to an agent, and neither is, *a fortiori*, that such an ethical disposition be voluntarily acquired. Indeed, no such argument for the voluntary acquisition of ethical dispositions is to be found in the *EE*.

Let me now say something about condition (ii), the lack of moral pain condition. As we shall see in Chapter 7, the lack of pain condition and its contrary, pain, involve rather sophisticated, focused attitudes towards the ethically salient features of one's practical situation. For instance, if an agent is not pained at going to the battle field, and if this lack of pain is going to be the relevant one, then it better be a lack of pain focused on a picture of the battle field *qua* e.g. the occasion for all sorts of courageous conduct, rather than *qua* the place where one might get wounded or even loose one's life. This sort of focused lack of pain and its contrary certainly require a certain level of ethical development, but I see no reason to think that children, or at any rate, children old enough to absorb the message conveyed by praising and blaming *logoi*, are in principle barred from fulfilling this condition (indeed, if they were, moral education would be impossible). Nor should incontinent agents be barred from having the sort of moral attitude essential to this focused lack of pain and its contrary, even though their incontinent conduct is not expressive of this (I argue for this in Chapter 2, 3 and 7).

What about condition (i) concerning voluntariness? It is not clear to me whether Aristotle ever managed to offer a satisfactory *positive* account of voluntariness. At any rate, his approach to these positive conditions is clearly negative, in the sense that it is arrived at *via* an analysis of the conditions of *in*voluntariness. This negative approach is present in the *EE* and the *Rhet.*, and is quite conspicuous in *EN* III 1.[12] According to the latter chapter, "*given that* the

---

[12] The *EE* for instance, says that acting through unawareness of the relevant circumstances renders one's action involuntary (1225b6-7) adding immediately: "and the opposite is voluntary (1225b7-8).

involuntary is <what is done> though violence or factual unawareness, one could think (*doxeien*) that voluntary is that action the source of which is in oneself, while one is aware of the particular circumstances in which it occurs" (*EN* 1111a22-24). This is why Aristotle does not concentrate on the conditions of voluntariness as much as he does on the conditions of *in*voluntariness, i.e. the conditions defeating ethical ascription.

Here I will follow the Nicomachean negative approach, but in a slightly more restricted way than Aristotle's own negative approach. I will of course be focussing on the conditions of involuntariness and pain condition, but Aristotle's own negative approach is also based on the assumption (which I have reasons to doubt, see Chapter 2, Sections D-E) that an analysis of these conditions is a good method for arriving at the *positive* conditions of voluntariness. The more restricted negative approach that I follow was justly praised by J. L. Austin, on the grounds that the notions of responsibility and freedom (the notions Austin is particularly interested in) are not names for a characteristic of actions, but rather concepts that, when applied to actions, "rule out the suggestion of some or all of its recognized antitheses" (i.e. rule out the suggestion that any factor that may render one's action un-free or such that one is not responsible for it, is at play). Because of this, Austin says that "'freedom' is not a name of a characteristic of actions, but the name of a dimension on which actions are assessed."[13] This is clearly true of Aristotle's notion of the voluntary (*to hekousion*). Let me explain this.

I do recognize that the notion of the voluntary plays *some* role in *De Motu Animalium* (703b3) where it is used in the rather physio-psychological context of a mechanical explanation of animal motion. Still, it is extremely significant to note that 'voluntary' (*hekousios*) as predicated of animal motion, and without reference to praise and blame, only occurs *once* in *De Motu* (and I think only once in the whole of the Aristotelian *corpus*) by contrast with the myriad instances in which this notion,

---

This negative notion of the voluntary is reflected in the *EE* definition of the voluntary as (partly) (i) what the agent does 'not in ignorance' (*mê agnoôn*, 1225b8-9). This same definition includes a second condition, namely that what the agent does *mê agnoôn*, he also does (ii) 'through his own agency' (*di' auton*), 1225b9; but it is clear that Aristotle has arrived at this notion of acting through one's own agency *via* his analysis of violence (*bia*) in *EE* II 8. The *Ars Rhetorica* makes this latter point explicit, by defining voluntary actions as "those things that we do with knowledge and *without violence* (*hosa eidotes kai mê anagkazomenoi*)" (1368b9-10).

[13] Austin, J. L. (1956/7), p. 8.

in its ethical use (i.e. as 'openness to assessment'), is found in the *EE*, the *EN* and the *Rhet*.[14] This is pretty strong evidence that the notion of voluntariness is primarily used, as predicated of an agent S (*hekôn*) or an action A (*hekousia*), to mark the fact that S or A is open to evaluation, i.e. open to praise or blame; and being 'open to praise and blame' here does not have to depend on a positive characterization of S or A in whatever terms (whether psychological or metaphysical) that may be thought to be the ultimate basis of such openness to evaluation.

If this is the case, and '*hekousion*' as predicated of an action or '*hekôn*' as predicated of an agent, signal the fact that these entities are open to assessment, rather than (primarily) standing for a positive property, whether psychological, physiological or metaphysical, then there is a sense in which merely indicating that none of the conditions rendering these entities *close* to assessment is present, is tantamount to characterizing them in terms of the voluntary. Furthermore, I take it that assessment in the sense that we are interested in, that is, the practice of attributing an ethically significant action A to a reason-responsive agent, is usually *not* consequent upon first *establishing* that A is voluntary because none of the conditions that render it close to assessment are present. This would be clearly unrealistic. Such attribution rather takes this voluntariness (or the absence of attribution-defeaters) *for granted*: it is then the task of the recipient of praise and blame, or the apt members of his community,[15] to come up with an explanation for why the attribution is unwarranted, and this will be an explanation in terms of the conditions of involuntariness. Accordingly, I will proceed on the supposition that the voluntariness of one's actions is initially assumed by one's spectators, and that such an assumption is nothing else above and beyond the assumption that a rebutting explanation in terms of the conditions of involuntariness is, in principle, lacking.

Notice that accounting for this assumption or presumption of voluntariness in these purely formal terms has the advantage of leaving open the question as to how exhaustive is the set of rebutting explanations available to the recipient of the

---

[14] This surely lends some support to Bernard Williams' contention that in Plato's and Aristotle's writings, "the functions of the mind, above all with regard to action, are defined in terms of categories that get their significance from ethics." Williams, B. (1993b), p. 42.

[15] That is, those who have access to the relevant facts about the agent. They are necessary, because no one is very eager to challenge the praise bestowed upon oneself.

attribution or the apt members of his community; for the assumption of voluntariness, accounted for in these purely formal terms, does not require this sort of exhaustiveness. I take it that this is an advantage, because the set of rebutting explanations available is clearly open to revision in the light of new scientific evidence, new philosophical arguments, or the new deliverances of common sense. It does demand, however, *some* account – and a good account – of the rebutting explanations traditionally available to the agent, and plausibly assumed by the spectators to be lacking when the attribution is being made. This is the account provided by Aristotle.

C. The defeaters of ethical ascription

In the second part of this dissertation we shall be concentrating our efforts on the factors defeating ethical ascription: the conditions of involuntariness and the pain condition. Of course, one may object that in his discussion of the conditions of involuntariness, Aristotle does not clearly separate the pain condition from the conditions of involuntariness *per se*, for he seems to imply that pain (i.e. the relevant sort of pain, at any rate) is necessary for involuntariness *per se* (e.g. *EN* 1110b18-24). This failure to keep involuntariness *per se* and pain separated, as I will suggest in Chapter 7, may be partly due to the fact that Aristotle does not always clearly distinguish between what makes an *agent* involuntary (*akôn*) vis-à-vis an involuntary action (i.e. the relevant pain), and what makes such an action involuntary (*akousion*) in the first place; but he does make this crucial distinction.

Now, the two main conditions that defeat ethical ascription are violence (Chapter 3) and factual error (Chapter 6). These conditions present their own problems of interpretation (to which I tend in these chapters), but a serious challenge to the negative approach arises if one takes these two conditions as furnishing a complete and fixed picture of what Aristotle has to say about the defeaters of ethical ascription (an interpretation that is encouraged by Aristotle himself): if this is the case, then one can argue that Aristotle does not have a good account of the (traditionally available) defeasibility conditions of ethical ascription. I can think of four such conditions or families of conditions, which may be thought to

be omitted from Aristotle's defeaters: (A) lack of assent; (B) causes of internal compulsion like extreme fear; (C) justifications (as different from excuses) and (D) the non-culpability of factual error or violence. Let me take these in turn.

(A) Even if S's behaviour *x* is involuntary, S's 'moral' aversion- or compliance-attitudes towards *x* (and perhaps also to *someone else's* ethically significant behaviour) can be an important basis for an ethical assessment of S, in so far as these attitudes express S's evaluative outlook. Here the presence a moral-aversion attitude in particular figures as a defeater – or so I contend in Chapter 7 – by 'backing up' the rebutting explanation offered by the agent or the apt spectator: e.g. 'he was forced to flee from the battle field (i.e. he is not thereby a coward), and he was clearly distressed, in the relevant way, by it (i.e. he is thereby a non-coward agent)' or 'he returned the deposit on time through factual error (i.e. he is not thereby just), and indeed he was annoyed that this happened (i.e. he is thereby non-just).'

(B) Not only is it the case that many motions that occur in our body and have an external source, thus occurring 'through violence' (*bia*[*i*]), enjoy nonetheless our assent. It is also the case that many actions that *are* indeed performed by the agent himself are also the product of (what we may think of as) *internal compulsion*. So, apparently Aristotle's two simple and apparently discrete conditions are also insufficient at this level. The sources of internal compulsion, in the Aristotelian sense, consist in those impulses that do not depend on the agent, in the sense that they are impulses that no good human specimen should be expected to resist. These impulses render the actions arising from them involuntary. In Chapters 4 and 5 I show that Aristotle deals with this group of cases in an interesting way that makes his notion of 'internal compulsion' quite different from (and perhaps superior to) the modern notion of psychological compulsion. Quite apart from internal compulsion in this sense, there are also conditions about the agent that make him an unsuitable recipient of praise and blame in a sustained manner, for instance his being a psychopath. The question is not whether Aristotle recognized these general conditions, for he did have ways of characterizing agents in abnormal psychological

conditions.[16] The question is whether Aristotle attempts to answer the question: 'What is it about an organism (as opposed to its pieces of behaviour) that makes it an appropriate recipient of praise and blame?' In Chapter 2, I suggest that Aristotle does have an answer to this question, in terms of reason-responsiveness (I do not discuss such general agent-based conditions like psychopathy, because their absence is presupposed by the notion of reason-responsiveness).

(C) Now, assuming that internal compulsion (in the sense captured by what I call 'objective overwhelming coercion') is acknowledged by Aristotle as one of the defeaters of ethical ascription, these conditions (i.e. violence, objective overwhelming coercion, and factual error) seem to amount to mere *excuses* or better, voluntariness-cancelling explanations, by showing that a (bad) action was performed involuntarily. But certainly one should also expect Aristotle to recognize the particular role of *justification* - not now as a kind of ascription-defeaters, but as a *blame*-defeater - i.e. reasons to believe that, although one *did* the (prima facie wrong) action *voluntarily*, one was *justified* in doing it, in such a way that the blame is cancelled. Aristotle, I submit, also acknowledged (again, in his own way) the role of justification in defeating ethical ascription, in the course of analysing the nature of coercion (*anagkê*) as a plea, as I try to show in Chapters 4 and 5.

(D) And lastly, once all these additional defeaters are adequately acknowledged, the worry may still be felt that *even* when an action *x* is shown to be *involuntary* (in virtue of any of these defeaters), *if* it was the agent's own fault that he put himself in a situation where factual unawareness, violence, etc. become operative, then he is to be blamed *in virtue of the very action* performed under these circumstances, so that, strictly speaking (and, as I will argue, contrary to what Aristotle suggests), none of these defeaters is a sufficient condition of involuntariness. I argue in Chapter 6 (Sections C-F) that this objection loses most of its force when one realizes that it is based on an assumption that can be denied (what I call 'the conservative assumption') and that could have been reasonably denied by Aristotle himself

---

[16] See for instance the *EN* VII 5 discussion of agent-based conditions like disease or madness.

If these four ways of 'upgrading' the conditions defeating ethical ascription are acknowledged, the resulting account of these conditions turns out to be surprisingly rich and appealing. If by the end of Chapter 7 the reader has come to appreciate these assets, I believe the purpose of the second part of my thesis has been fulfilled.

PART ONE

TOWARDS AN ACCOUNT OF ETHICAL ASCRIPTION

I

AGAINST THE STRAWSONIAN INTERPRETATION

THE NEGATIVE ARGUMENT

In both the *EE* and the *EN* Aristotle develops a theory of voluntariness, and suggests that such a theory is relevant for his overall ethical project, because we are praised and blamed for our voluntary responses. In this chapter I argue against an influential interpretation of Aristotle's concern with the voluntary/involuntary distinction in his Ethics, which is based on an account of this connection between voluntariness and praise and blame. I call this the 'Strawsonian interpretation', because, as I show in Section B, it relies on Peter Strawson's conception of moral responsibility.

A. Voluntariness, praise and blame.

In *Ethica Eudemia* II 6, where Aristotle first introduces the voluntary/involuntary distinction (*hekousion/akousion*), he employs four expressions (perhaps with the same reference but different senses) that somehow correspond to a notion of causal responsibility. He says that (i) the human being alone amongst animals is the *source* (*archê*) of certain actions (1222b19, b29; 1223a3, a16, a17); (ii) the human being is a *master* (*kurios*) presumably of the same actions of which he is said to be a source of (1223a5, a7); (iii) these actions of which he is a master and the source depend on him (*eph' hautô*[i]) (1223a7); (iv) and finally he claims – in what seems to be a terminological digression aimed at defining these concepts in terms of each other – that the human being is *aitios* for/of these same actions (1223a8, 15, 16). This latter is obviously connected to what could be considered a fifth claim: (v) that the human being is the *cause* (*aitia*, *aition einai*) of his actions (1222b31; 1223a19). These notions are also present, although less conspicuously, in *EN* III.[1]

Aristotle apparently thought that any of these notions, including of course the notion of voluntariness (*to hekousion*), could be substituted for any other *salva veritate*.

---

[1] *EN* 1110a1, a15-16, a23, 1113b18, b20, 1114a20 for *archê*; 1113b24, b30, b34, 1114a4, 1114b3 for *aitios*, *aition*; 1113b32, 1114a3, a32, b30 for *kurios*; *EN* III 3 for the notion of *eph' hautô*[i]).

EE II makes this connection in various conceptual steps: (i) "it is clear that all the actions of which a human being is the source (*archê*) and the master (*kurios*) may either happen or not happen, and that it depends on him (*eph' hautô*[i]) whether they happen or not, as he is the master (*kurios*) of their existence or non-existence" (1223a4-7); (ii) "of things which it depends on him to do or not to do he is himself *aitios*, and what he is *aitios* of depends on him" (1223a7-9); (iii) "we all agree that each man is the cause (*aition*) of all those things which are voluntary (*hekousia*) … and that he is not himself the cause of all those things that are involuntary (*akousia*)" (1223a17-20). And the *EN* claims that "actions that have their source (*archê*) in us, they themselves, too, depend on us (*eph' hêmin*) and are voluntary (*hekousia*)" (1113b20-21).[2]

Aristotle explicitly says that a source (*archê*) of the becoming of *x* is always a cause (*aitia*) of *x*, indeed the *first cause* (*EE* 1222b31). Hence, when Aristotle says that "excellence and vice have to do with those actions of which one is oneself *aitios and the source*" (1223a15-16), '*aitios*' (the adjective) seems to convey more than the idea of cause (*aitia*, as a noun), which is already involved in the notion of source (*archê*). It is then quite tempting to translate '*aitios* plus genitive' as '*responsible* for x' rather than as 'cause of x', as translators tend to do.[3] Indeed, if there is a word in Greek, more or less equivalent to the English 'responsible', that word is the adjective '*aitios*', the core meanings of which are 'responsible', 'culpable', or 'guilty' (*ho aitios* was indeed the culprit in legal contexts, and *hê aitia* the accusation). The major problem with translating '*aitios*' as 'responsible' in these contexts is that the English term bears strong nuances of *accountability* or *answerability*,[4] a concept soon to be elucidated, and the question that I will be addressing in this chapter is precisely whether Aristotle's concern in his Ethics with these notions of being the source, the master and *aitios*, of a given response *x*, or of *x* depending on him or being voluntary, reflects a concern with accountability or answerability. To avoid prejudging the nature of Aristotle's concern with these notions in these contexts, I will talk about *x* being *ascribed* or *attributed* to someone, instead of someone being *responsible*

---

[2] Cf. *EN* 1136b27-28.
[3] Examples are H. Rackham, M. Woods and J. Solomon.
[4] As a matter of fact, this is the meaning of the Latin '*responsus*' (from which the English 'responsible' ultimately derives), the participle of '*respondere*', 'to answer'.

for *x* (indeed, I will leave '*aitios*' untranslated), and about *ascribability* or *attributability* instead of *responsibility*. I will do this unless I am reporting or referring to particular views about the nature of Aristotle's concern with these notions.

Now, for the purposes of the present chapter, the crucial step in Aristotle's argument is the connection of these causal notions of S being the source, the master, or *aitios* of a given response *x*, or of *x* depending on S or being voluntary (*hekousion*), with praise (*epainos*) and blame (*psogos*). Aristotle introduces this crucial connection thus:

And since virtue and vice and the operations that spring from them (*ta ap' autôn erga*) are respectively praised and blamed - for we do not praise and blame for what is due to necessity, or chance, or nature, but only for what we ourselves are *aitioi* of; for what another is *aitios* of, for that he bears the blame or praise - it is clear that virtue and vice have to do (*peri*) with these things, i.e. those actions of which a human being is himself *aitios* and source. *EE* 1223a9-15[5]

The opening paragraph of *EN* III states a very similar reason for pausing over the causal notions just referred to, but in terms of the notion of voluntariness (*hekousion*):

Since virtue is concerned with (*peri*) affections and actions, and on the voluntary ones praise and blame are bestowed, whereas on involuntary ones sympathy, and sometimes even pity, to distinguish the voluntary and the involuntary is presumably necessary for those inquiring into the subject of virtue. *EN* 1109b30-34

The reason why voluntariness is a topic of significance for the Ethics, according to the *EE* and the *EN*, can be summarized thus: Given that Ethics is an enquiry about the highest good for human beings *qua* such (what that good is and how to achieve it), and given that the highest good is human flourishing (*eudaimonia*) and that this consists primarily in virtuous activity, then any serious ethical enquiry has to engage in a discussion about *the conditions of virtuous activity*. This is the implicit background of the passages. In both treatises[6] Aristotle says that, (1) since virtue and vice and the responses virtue and vice

---

[5] Solomon's translation with slight modifications. In general, if I don't refer to the translation I am using, this is because the translation is mine.
[6] See also *MM* 1187a20-21.

are *concerned with* (*peri*) – i.e. virtuous activity – are praiseworthy and blameworthy respectively (implied by the *EN* 1109b30-34 passage), and (2) since praise and blame are bestowed upon voluntary responses, responses of which we are *aitioi*, then (3) we must study the voluntary.

Now, 'affections and actions that are the *concern* of virtue and vice' (premise (1)) is ambiguous between (A) the kind of responses by which excellences and faults of character *come about*, which involve voluntary actions (*EN* 1113b26-27; 1114a12-16); and (B) the kind of responses that ethical dispositions *dispose us to do*. This ambiguity of 'virtuous activity' between (A) and (B) is promoted by Aristotle's belief that the *actions* in both sides of the continuum of character acquisition are *of the same in kind*: "the kind of things by which excellences come about are the kind of things they dispose us to do [lit. are the things of which they are 'practitioners', *praktikai*], in accordance with themselves (*kat' autas*)" (1114b26-27)

The crucial claim, that on the basis of which many interpreters have tried to resolve the aforementioned ambiguity, is Aristotle's claim that one way of learning what the conditions of virtuous activity are is by pointing at our (or the Athenian, or Aristotelian) practices of praising and blaming. As Aristotle says, a mark of virtuous activity is that it is praised, and a mark of faulty activity is that it is blamed (this is compatible with the two senses of 'virtuous activity' just distinguished), and this praising and blaming presuppose that they are activities of which we are *aitioi* (i.e. activities that are attributed *to us*), for when we learn that they have been the result of necessity, chance, etc. praise and blame are withheld. Now, this connection between voluntary responses, those responses we are *aitioi* of or that are 'ascribed to us', and praise and blame, is crucial for understanding the role played by Aristotle's theory of the voluntary in his Ethics: on the basis of certain modern accounts about the nature of praise and blame, some interpreters (whom I call the 'Strawsonian interpreters') have argued that Aristotle, in his discussions of voluntariness in the Ethics, is *exclusively* concerned with virtuous activity of type (B), that is, the type of activity that virtue disposes us, mature ethical agents, to do; and that Aristotle's theory of voluntariness in his Ethics amounts to a theory of moral

responsibility, at least in the sense in which modern philosophers in the Strawsonian tradition have understood this latter concept. In the next section I expound the Strawsonian interpretation, which will be the chief concern of this chapter.

B. The Strawson-inspired interpretation of Aristotle

In modern philosophy there is a well-defined theory of moral responsibility that finds its classic expression in P. H. Strawson's paper 'Freedom and Resentment'.[7] By way of preliminary, we can say that according to this view, S is morally responsible for an item of behaviour *x* (or the consequences of *x*) if and only if it is *appropriate* for the members of S's moral community to react to *x* by feeling resentment (if they are the affected parties) or moral indignation (if someone else is), towards S for having done *x*, if *x* is an evil; or by feeling gratitude (if they are the beneficiaries) or moral respect (if someone else is) towards S for having done *x*, if *x* is a good. These reactions are what Strawson calls 'other-reactive attitudes'. They can be personal, when defined from the perspective of the affected or benefited party, as resentment and gratitude are; or impersonal or vicarious, when viewed from the perspective of a party who is not directly affected or benefited, but who nonetheless is entitled to feel resentment or gratitude *on behalf* of the directly interested party.

According to Strawson, it is appropriate to hold these attitudes towards S for having done *x* only if three conditions are met; about S, about his item of behaviour *x*, and about the conditions under which *x* has been performed: (i) S must himself be a member of the moral community; (ii) *x* must be such that it breaches (if the attitudes are negative) or complies with (if they are positive) what Strawson described as "an expectation of, and demand for, the manifestation of a certain degree of goodwill or regard on the part of other human beings"[8] towards ourselves and other human beings; and (iii) *x* must be voluntary (or whatever concept captures those conditions that make it appropriate to hold the reactive attitudes towards S *for causing x*).

---

[7] For a detailed exposition of the Strawsonian model of moral responsibility see Wallace, J. (1994).
[8] Strawson, P. F. (1963), p. 84.

Some important details of the Strawsonian story will be filled in later in this section and in Section F. Let me now mention three fundamental points about these three conditions, by way of preliminary only. These three points are essential to the aim of this section, which is to suggest that the Strawsonian view has inspired a particular line of interpretation about Aristotle's concerns with voluntariness and related causal notions in his Ethics. The first point is about condition (i): *S* is a member of the moral community, only if S is himself capable of holding the reactive attitudes mentioned above, along with the capacities necessary for this (e.g. the capacity to put oneself in others' shoes). Being capable of the reactive attitudes involves capacity not only for holding 'other-reactive' attitudes towards others, but also for self-reactive ones like guilt, compunction or remorse: the latter are the attitudes that S himself would feel when being the target of the other-reactive attitudes, or even when considering that these other-reactive attitudes towards himself would be justified.[9] Because they lack these capacities, children and psychopaths are denied membership in the moral community, according to Strawson.[10] The second point is about (ii): Strawson thought that this demand of goodwill or regard that we make on other human beings is an *essential* element of our inter-personal relationships and our moral life as we know it, and it is what ultimately grounds the reactive attitudes: these are attitudes that we hold towards S whenever S's actions have breached or met this basic moral demand. Because of this, it is correct to say that holding the reactive attitudes towards S for his having done *x* are ways of *holding S accountable* or *answerable* for having breached or met this basic moral expectation,[11] also alternatively called by Strawson 'moral obligation'.[12] The last point is about condition (iii): this condition simply captures the fact that S's good will or regard for others is only expressed in that

---

[9] Thus, the readiness on the part of S to *acquiesce* in the infliction involved in the other-reactive attitudes towards himself (when these are negative) as 'his due' or as 'just' (i.e. without resentment), is also a requirement for regarding S as a member of the moral community (*ibid*. p. 91).

[10] So Strawson says that an agent is excluded from the moral community, that is, from the domain of other-regarding reactive attitudes, if it "is seen as excluded from ordinary adult human relationships by deep-rooted psychological abnormality – or simply by being a child" (*ibid*, p. 81). Considerations of this sort (e.g. 'he is just a child', 'he is a psychopath'), says Strawson, "make us see the agent as other than a morally responsible agent" (*ibid.* p. 85).

[11] Although Strawson himself does not describe his account in this way.

[12] *Ibid*. p. 86

behaviour which is voluntary, and that, as a corollary of this, the reactive attitudes towards S for his having done *x* are inhibited whenever *x* is involuntary.

Now, because Aristotle holds that the question of praising an agent for an action *x* can arise if and only if *x* is voluntary (*EE* 1228a10-12, *EN* 1109b31) – corresponding to condition (iii) *supra* – and because of the post-strawsonian temptation to regard Aristotelian praise and blame as S-reactive attitudes[13] – corresponding to condition (ii) *supra* – and to find in Aristotle's Ethics criteria that appropriately distinguish moral from non-moral agents in the Strawsonian sense – corresponding to condition (i) *supra*; it is possible to think that, according to Aristotle, for S to voluntarily do a blameworthy or praiseworthy action *x* and to be blamed or praised accordingly, is for S to be an appropriate target, on the occasion of his having done *x*, of S-reactive attitudes. This Strawsonian analysis of Aristotle's assertions to the effect that S is blameworthy or praiseworthy for having voluntarily done *x* is, I shall argue, on the wrong track. Let us take a closer look at the main features of the Strawsonian interpretation.

(I) <u>*Aristotelian praise and blame as S-reactive attitudes*</u>. Terence Irwin, for instance, although often using the general term 'responsibility' and referring to 'Aristotle's theory of responsibility', makes clear that by these terms he means the theory of what it means to 'hold someone responsible for his actions': "It is fair to expect a good theory of responsibility" he says, "to explain and justify our treatment of someone we *hold responsible* for his actions",[14] and he makes it clear that for him 'holding people responsible' corresponds to this very notion of responsibility à la Strawson:

Aristotle says *praise and blame are the proper reactions to responsible agents*; this will do as a rough summary, if we include *gratitude, resentment, thanks, indignation, admiration, condemnation*, with simple praise and blame... being responsible [for an action] is *being a reasonable candidate for these attitudes, the sort of agent doing the sort of action for which praise and blame are normally justified*.[15]

---

[13] The phrase 'reactive attitude' is too vague: since any response to anything is a reaction, practically any attitude can be described as 'reactive'. I will use 'S-reactive attitude' (meaning 'Strawsonian reactive attitude') to designate the moral sentiments on which Strawson himself focuses, and which are central to his conception of moral responsibility.

[14] Irwin, T. (1980), pp. 133-4.

[15] *Ibid*. p. 134. My emphasis.

The fact that Irwin includes gratitude, resentment, indignation and condemnation (which he calls 'attitudes'), shows that he is almost certainly thinking in Strawson's terms, and his including praise and blame in the list shows that he is interpreting Aristotelian praise and blame as S-reactive attitudes. This identification is also evident in Sauvé-Meyer's book *Aristotle on Moral Responsibility.* Sauvé-Meyer believes that Aristotle's account of voluntariness at *EE* II 6-9 and *EN* III 3 "is intended to capture the conditions of moral responsibility for action",[16] and the conception of moral responsibility she endorses is roughly the Strawsonian one. According to Sauvé-Meyer, Aristotle identifies the conditions of voluntariness (or of being *aitios* and the source of one's actions) with the conditions of praise and blame, and praise and blame are S-reactive attitudes:

Praise and blame are simply two of a wide range of *interpersonal attitudes* that are of central importance to our lives as moral agents. An agent who does not *merit* praise or blame for her actions is also an inappropriate recipient of such attitudes as gratitude, resentment, forgiveness, anger, friendship, and love. These attitudes are *based on certain expectations of interpersonal regard and they express our reactions to the fulfillment or non-fulfillment of those expectations*. To be subject to such reactive attitudes is to be subject to those expectations, and to be subject to these expectations is to be subject to the *demands of morality*, *and hence to be a moral agent*. On this broader construal of moral responsibility, moral responsibility is the property of an agent that makes her subject to the demands of morality, and hence subject to moral evaluation in the light of these demands.[17]

Accordingly, a feature of what I call the 'Strawsonian interpretation' of Aristotle's concern with the conditions of voluntariness in his Ethics, is that Aristotelian praise and blame are S-reactive attitudes.

(II) *Aristotelian recipients of praise and blame as accountable agents.* A second feature of the Strawsonian interpretation follows immediately from the first one, and it is already hinted at in the two passages from Irwin and Sauvé-Meyer just quoted. If these authors' interpretation is in fact inspired by Strawson's conception of moral responsibility,

---

[16] Sauvé-Meyer, S. (1993), p. 6.
[17] *Ibid*. p. 18. My emphasis.

then it must somehow cope with Aristotle's claim that animals and children are also capable of voluntary action (*EN* 1111a25-6), for neither animals nor children are, as Strawson observes, the appropriate recipients of moral indignation, resentment, or genuine gratitude, nor do they have the competences to hold genuine S-reactive attitudes towards others: they are not members of the moral community, nor, as a consequence, morally accountable or answerable agents. Accordingly, the Strawsonian interpreter needs to provide the Aristotelian equivalent of Strawson's condition (i) in p. 24*.* It is unsurprising, therefore, that Irwin characterizes (in the passage just quoted) the 'responsible agent' in Strawsonian terms, as "a reasonable candidate for these attitudes, the sort of agent doing the sort of action for which praise and blame are normally *justified*" and suggests that Aristotle himself justifies "our treatment of someone we *hold responsible* for his actions" by identifying "responsible agents with *agents capable of effective decision*" or prohairetic agents (i.e. agents capable of effective *prohairesis*, rational choice),[18] for neither children, nor of course animals, are capable of *prohairesis* (*EN* 1105a31).[19] Nor should it come as a surprise, likewise, that Sauvé-Meyer deems it necessary to find in Aristotle "criteria that underlie our discriminations between agents who are morally responsible, and hence *full-fledged members of the moral community*, and those who are not".[20] According to Sauvé-Meyer, Aristotle provides such criteria in his analysis of what it is for an agent to be in possession of an ethical disposition,

---

[18] Irwin, T. (1980), p.134. My emphasis.

[19] The main reason why Irwin (1980) thinks it is necessary to amend what he calls Aristotle's 'simple' theory of responsibility (the theory merely based on the conditions of voluntariness), is that (i) "animals and children are not responsible for their actions". This premise, if added to the claim that (ii) "A is responsible (a proper candidate for praise and blame) for doing x if and only if A does x voluntarily", and (iii) "Animals and children act voluntarily", results in a contradiction (see Irwin, T. (1980), p. 125). I think that Irwin makes a double mistake here. One is to include animals in Aristotle's theory of responsibility. It is true that Aristotle says that animals act voluntarily (*EN* 1111b6-9), but why conclude from this that they are candidates for praise and blame? After all, there is, as we have noticed in the Introduction, a physio-psychological use of '*hekousion*' available to Aristotle, which is not connected to praise and blame. The case of children of course, is different, as we shall see. Irwin's second mistake is to assume without argument that (granting the point about animals) children are not responsible for their actions. Irwin wants to say that "Aristotle clearly assumes that... children are not subject to legal and moral sanctions" (*ibid*. p. 125). This is an interesting claim. Perhaps in Aristotle's world children were not subject to the legal sanctions applicable to adults, but textual evidence is needed to extend this to *moral* sanctions. Furthermore, even if direct or indirect evidence can be found to support this latter claim, not all the blame that is relevant to the ascription of ethically significant behaviour to agents is expressed in the form of *sanction*, or so I will argue. Blame equals moral sanction from the view point of accountability, but not, as we shall see, from the point of view of attributability.

[20] Sauvé-Meyer, S (1993), p. 19. My emphasis.

characterized by him as a *hexis prohairetikê* (*EN* 1106b36). For one of Aristotle's requirements for virtue and vice is the capacity of reason, and in particular, *prohairesis*: rational choice. According to Sauvé-Meyer, since animals and children do not *have* such capacity of rational choice, Aristotle thereby provides us with criteria to exclude them from the category of (Strawsonian) moral agents:[21] animals, psychopaths, and children are all excluded from this category *for the same reason*.[22] It is obvious to me that these theorists are guided by the question of how and whether Aristotle characterizes the 'morally responsible agent' because they approach the question about Aristotle's concern with the conditions of voluntariness in his Ethics, from a Strawsonian perspective, and take Aristotle's concern with these conditions to be a concern with moral responsibility à la Strawson.

(**III**) *Aristotelian praise and blame as non-prospective attitudes*. The last feature of the Strawsonian interpretation that I will be dealing with is implicit in the other two features, and it is thus particularly crucial to the arguments I will offer against it. Strawson argued that general considerations about the condition of an *agent* such as 'he is just a child' or 'he is insane' (as distinct from considerations about *particular actions*) inhibit the S-reactive attitudes in a generalized way, i.e. they encourage us to view any of these agents as the *sort of agent* who is not to be held accountable for complying with the demands of morality (as distinct from someone who is not to be held accountable for this or that particular action). But Strawson also thought – and this is better regarded as a different thesis of his – that these general, agent-based considerations promote instead an altogether different sort of attitude that he calls the 'objective' attitude, a sort of attitude that is, he thought, profoundly *opposed* to S-reactive attitudes:

To adopt the objective attitude to another human being is to see him, perhaps, as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be

---

[21] The most significant difference between Irwin and Sauvé-Meyer in this respect is that she takes 'virtue' and 'vice' in a wide sense, which includes both continence and incontinence (see *EN* 1150b32; *EE* 1223a37, b12) – *ibid.* p. 29. Irwin's 'agents capable of effective decision', on the other hand, seems to rule out incontinent agents, for incontinent agents are capable of decision, but not of *effective* decision (although in Irwin, T. (2007), p. 180 he confines praise and blame to "agents capable of election".)
[22] Sauvé-Meyer, S. (1993), p. 23.

taken account, perhaps precautionary account, of; to be managed or handled or cured or trained; perhaps simply to be avoided... [23]

It seems then that anyone meaning to attribute to Aristotle a theory of moral responsibility à la Strawson is committed to arguing that the sort of praise and blame Aristotle thought could only be bestowed upon voluntary actions, does not correspond to merely 'objective attitudes', as these are defined by Strawson. It is therefore unsurprising to find Irwin arguing that Aristotelian praise and blame are *not* 'objective' attitudes:

Now why should we follow Aristotle in associating praise and blame with the capacity for deliberation and decision? At first sight the connection is not obvious. We may change the behaviour of some animals, and certainly of children, by praise and blame; and in general, vigorous praise and blame may affect someone's behaviour whether or not he reflects or deliberates about it. *But this is not our normal attitude to responsible agents*. When we react with praise, blame, resentment, gratitude, and so on, we do not normally mean to cause some irresistibly compulsive desire in the agent to do what we approve of; we do not normally expect that he will be incapable of resisting his reformed desires even if he things it better not to act on them.[24]

Nor is it surprising to find that Sauvé-Meyer regards as essential the task of showing that Aristotelian praise and blame do not correspond to the 'objective' attitudes:

If the praise and blame in question are retrospective moral evaluations of agents for what they have done or failed to do, then we have good reason to suppose moral responsibility is Aristotle's concern in these contexts. However ... the praise and blame for which Aristotle thinks voluntariness is necessary might simply be *tools of behavioural control and character formation justified by purely prospective considerations*. If this is the case, then his concern with voluntariness is not a concern with moral responsibility.[25]

Notice that the Strawsonian contrast between 'reactive' and 'objective' attitudes is now framed, by Sauvé-Meyer, in terms of retrospective and prospective ways of justifying praise and blame. This way of framing the contrast between S-reactive and objective

---

[23] Strawson, P. F (1962), p. 79.
[24] Irwin, T (1980), p. 134. My emphasis. See also Irwin, T. (2007), p. 178.
[25] Sauvé-Meyer, S. (1993), p. 40. Italics are mine.

attitudes is thoroughly warranted by Strawson's own view, but it is misleading. As it is clear from Protagoras' celebrated theory of prospective punishment (Plato's *Protagoras* 324a-c.), prospectively praising and blaming an agent S for an action *A* are practices that seek (1) to encourage S to perform, or to deter S from performing, actions relevantly similar to *A* in the future, or, alternatively, (2) to increase or decrease the probability of the performance of type-*A* actions as performed by others. Another clear prospective function, not distinguished by Protagoras, is (3) to let others know, for the purposes of their future policy towards S, and dealings with S, that S is someone who does actions of type *A,* e.g. to warn prospective employers. I think this definition captures more or less accurately the essence of forward-looking or prospective praising and blaming.[26] It does not capture, however, the essence of 'objective' or (what I prefer to call) 'instrumentalist' attitudes. This is because not all forward-looking attitudes are instrumentalist or 'objective'. Strawson thought that the adjective 'moral' should be withdrawn from the latter, because the 'objective' attitudes 'instrumentalize' their targets: so Strawson speaks of the 'objective' attitudes as being efficacious "in regulating behaviour *in socially desirable ways*", and as "instruments of policy" that offend "the humanity of the offender himself".[27] I will argue in the next section that many cases falling under (1), however, do not satisfy this description in that they are not justified by purely instrumentalist considerations. And it can be noticed, straight away, that there are cases falling under (3) that do not either; e.g. in warning a prospective employer about S I am not necessarily treating S in an 'instrumentalist' way.

It is important to note that Strawson tends to identify 'objective' attitudes with *prospective* ones, but our subsequent discussion will greatly benefit if (i) we refer to Strawson's 'objective' attitudes as 'instrumentalist', for it is the fact that they are purely

---

[26] One can argue that praise and blame become efficacious social tools of behavioural manipulation only when targeted at actions that can be ascribed to the agent, so that actions of type A are by definition voluntary. So Smart, J. J. C. (1961), p. 68: "Suppose Tommy at school does not do his homework. If the schoolmaster thinks that this is because Tommy is really very stupid, then it is silly of him to abuse Tommy, to can him or to threaten him. This would be sensible only if it were the case that this sort of treatment made stupid boys intelligent. With the possible exception of certain nineteenth-century schoolmasters, no one has believed this."

[27] (That is, if the offender is a member of the moral community). Strawson, P. F. (1962), p. 89. Italics are mine. For Sauvé's explicit endorsement of this view, see Sauvé-Meyer, S. (1993), pp. 18-20.

justified by considerations of social utility, that grounds our unwillingness to classify them as 'moral' (i.e. 'moral' praise and blame); and if (ii) we bear in mind that 'prospective' or 'forward-looking' is much broader than 'instrumentalist'.[28]

The reason why this third feature of the Strawsonian interpretation (i.e. III) is so essential is that, as Strawsonians (wrongly) see it, if a *prospective* understanding of the nature of Aristotelian praise and blame is right, then their interpretation is wrong. Moreover, the prospective interpretation can offer a perfectly plausible account of why, according to Aristotle, children and animals act voluntarily, and of why, in Jean Robert's words, "Aristotle does not make more of the morally important differences between adults and children in this context."[29] Roberts' own view – let us call it the 'prospective conditioning view' - is precisely that *prospective* praise and blame[30] can account for this "curious fact" (her words):

Humans (or animals) are responsible for their voluntary actions in the sense that their particular desires are at least a crucial part of the explanation of what happens. Because voluntary actions thus reflect the variable and changeable aspects of souls, it is appropriate to praise and blame them. These are the sorts of actions that, in principle, *one can be trained or become habituated or be educated to do or not do*. Praise and blame, punishment and reward, are part of their training and education.[31]

Notice that Roberts includes a variety of attitudes like training, habituation, and education (this latter being prospective, but not instrumentalist), in her account of Aristotelian praise and blame. From this account, Roberts concludes that Aristotle's notion of voluntariness (as it occurs in his Ethics) "does not coincide ... with any later notion of

---

[28] For quite different reasons, the labels 'forward-looking' and 'backward-looking' are misleading, in that they suggest that it is a distinctive mark of prospective praise and blame that at least part of the aim of their being uttered is to prompt some sort of *response* from the agent targeted. But this is clearly false: it is clear that part of the utterance of blame-sentences for Strawson, and for anyone who thinks of blame as genuinely retrospective, seeks to get the agent to respond in certain ways, e.g. to apologize, to regret, to make an indemnification, or at least to defend himself against the charge.

[29] Roberts, J. (1989), p. 26.

[30] Which Roberts seems to equate to punishment or reward, see *EN* 1109b34-35.

[31] *Ibid*. p. 25. Further on this: "It is far from obvious, however, that Aristotle sees the actions of adults, in general, as different from those of children in a way that could be expected to infect his conception of the role or nature of punishment (or reward) in the two cases. One 'deserves' punishment, for Aristotle, if there is something wrong with one's soul of the sort that might be correctable. Thus animals, small children, and fully mature adults will all be in need of punishment for the same reason." (*Ibid*. p. 23).

moral responsibility. I take it that whatever precisely moral responsibility is, animals and small children do not have it".[32] This is not surprising, given the Strawsonian tendency to include *all* prospective attitudes under the category of 'objective'. Led by the same tendency, the Strawsonian interpreter, on the other hand, deems it necessary to firmly oppose the prospective conditioning interpretation.

The most elaborate attempt in this direction is Sauvé-Meyer's. Her attempt is to show that the sort of praise and blame that in Aristotle's view is only bestowed upon voluntary actions, is *not* the sort of praise and blame bestowed upon immature agents for their voluntary actions leading to the *formation* of their ethical dispositions (i.e. the sort of praise and blame involved in moral training), but rather a sort of praise and blame *exclusively* bestowed upon agents *already* in possession of those ethical dispositions, and bestowed upon them on the occasion of those actions of which only they, as possessors of ethical dispositions, are *reliable producers*. This is precisely Sauvé-Meyer's view, based on an analysis of certain passages (*EN* 1101b14-16, 1101b31-32; *EE* 1219b8-9; *MM* 1183b26-27).[33] Sauvé-Meyer concludes:

Aristotle intends his account of voluntariness to capture the conditions in which an agent's states of character are praiseworthy and blameworthy. And these conditions … are those in which an agent is *causally responsible for particular good and bad actions or feelings*. The sort of praise and blame the agent *deserves* for (epi + dative) these actions and feelings is therefore based on a *retrospective evaluation – for the question is whether the agent's character produced them*. Furthermore, they are *moral attitudes* – for they are essentially evaluative of the agent's states of character … Aristotle introduces a discussion of voluntariness into [his] general account of virtue and vice in order *to identify the causal conditions in which an agent's action may properly be taken to manifest a particular virtue or vice of character* – to give a criterion for deciding when, for instance, an agent's action is a case of acting the coward (*deilainein*), or acting intemperately (*akolastainein*), or in general acting justly (*dikaiopragein*) or doing injustice (*adikein*). When an agent acts the coward or acts intemperately, he is properly reproached as cowardly or intemperate, and when he does injustice he is properly praised as just. These evaluations are recognizably moral evaluations.[34]

---

[32] Roberts, J. (1989), p. 25.
[33] For my analysis of these passages, see Chapter 2, Section C.
[34] Sauvé-Meyer, S. (1993), p. 51.

In saying that an agent who acts the coward or acts intemperately is *morally* evaluated, Sauvé-Meyer is not employing the adjective 'moral' to translate the Greek '*êthikos*'. This would render her claim a mere tautology. She is rather making a substantive claim about the sort of responsibility that Aristotle claims human beings have with regard to their actions.

Now, the idea that Aristotelian praise and blame is *retrospective*, and therefore that it and the sort of responsibility derived from it is of a *moral nature*, is of course rooted in the (missleading) Strawsonian idea that 'moral evaluations' (to use Sauvé-Meyer's phrase) cannot be prospective, because 'prospective' equals 'instrumentalist' and 'instrumentalist' entails 'non-moral' (indeed, potentially 'immoral'). All genuinely moral evaluations are then non-prospective, that is, retrospective. We have then identified one crucial source from which the adjective 'moral' gets into the picture, according to Strawson:

(a) If (Aristotelian) praise and blame are retrospective (in the way the S-reactive attitudes are) then they are the sort of attitudes that acknowledge their recipient as a moral agent qua such, or perhaps as an end in itself, as opposed to regarding him as a mere *means* or 'instrument' (to use a Kantian dichotomy), as instrumentalist (or 'objective') attitudes do. As noted above, Strawson certainly thought that the adjective 'moral' should not be predicated of the instrumentalist ('objective') attitudes, because they 'instrumentalize' their targets: so he speaks of the 'objective' attitudes as being efficacious "in *regulating* behaviour *in socially desirable ways*", and as "instruments of policy" that offend "the humanity of the offender himself".[35] The operative words here are 'regulating' and 'instruments'. Strawson's view is surely that it is morally objectionable to treat moral agents *just* from the point of view of what is desirable for the wider society, and thus to 'regulate' or 'condition' their behaviour with this end in view. Let me call this 'the instrumentality argument'.

---

[35] Strawson, P. F. (1962), p. 89. My emphasis.

But there is also another source of the adjective 'moral' in 'moral evaluation', which is important to bear in mind:

(b) If (Aristotelian) praise and blame are retrospective (in the way the S-reactive attitudes are) then they are attitudes that *look back* to the agent as someone who 'deserves' or 'merits' such an attitude in virtue of an action of his that has breached or fulfilled a *moral demand* (e.g. Strawson's basic moral demand of goodwill).

The crucial point is that Sauvé-Meyer's interpretation is based on a sense of 'backward-looking' or 'retrospective' that does not seem to be, in any obvious way, relevant to the 'moral' status of praise and blame, in so far as it is not clear how it relates to (a) and/or (b). Sauvé-Meyer emphasises the *causal antecedent* of the action in the agent's character – and this is why (according to her) her account of Aristotelian praise and blame is *retrospective*. But how is this causal sense of 'retrospective' suppose to help us understand why calling an adult agent in possession of an ethical disposition 'just' or 'cowardly' for being the voluntary cause (*qua* having such disposition) of a just or cowardly outcome, is a distinctively *moral* evaluation, *in a non-tautological sense*, i.e. in a substantive sense that does not merely translate '*êthikos*'?

Perhaps the Strawsonian interpreter is relying on (a): if Aristotelian praise and blame are retrospective (in the way the S-reactive attitudes are) then they are the sort of attitudes that acknowledge their recipient as a moral agent in his own right, or perhaps as an end in itself, as opposed to regarding him as a mere means or 'instrument', as the 'objective' attitude does. I argue in the next section, Section C, that, (i) because the Strawsonian interpreter, following Strawson, identifies 'prospective' with 'instrumentalist', (ii) he rules out *all* prospective attitudes as non-moral; Since (iii) educational attitudes are indeed prospective attitudes, but (iv) educational attitudes *do not* involve an 'intrumentalization' of pupils, feature (a) fails to provide an adequate basis for regarding retrospective praise and blame *eo ipso* exclusively moral. Or perhaps the

Strawsonian interpreter is relying on (b): perhaps Aristotelian praise and blame are attitudes that *look back* to the agent as someone who 'deserves' or 'merits' such an attitude in virtue of an action of his that has breached or fulfilled a *moral demand* (in a non-tautological sense!). I argue in Section D that this is certainly not true of Aristotelian praise and blame.

Furthermore, because the three features of the Strawsonian interpretation mentioned in this section are so closely connected to each other, the arguments in Sections C and D inevitably cast doubt on the other two features of the Strawsonian interpretation (i.e. the view of Aristotelian praise and blame as S-reactive attitudes and the exclusive focus on the morally accountable agent).

C. Praise and blame as involved in moral education

As a way of introducing the argument of this section, let me first say something about the texts of Aristotle's Ethics on which Sauvé-Meyer bases her contention (i.e. feature II, p. 26) that, according to Aristotle, children are excluded from the category of moral agents *for the same reasons* as animals and psychopaths are: they do not have "the capacity of reason". The first relevant passage (quoted by her) is *EN* 1149b31-1150a1:

We do not call beasts either temperate or intemperate, except metaphorically when one kind of animal differs from another in violence and wantonness and in being ravenous. For they do not have decision (*prohairesis*) or reasoning (*logismos*), but rather are outside of <this> nature just like the mad among human beings.[36]

The fact that Aristotle does not mention children in this passage seems to me of great significance. Children are not 'outside of rational nature' in the way animals and madmen are. They lack reason only in the sense in which they are potentially, and not actually, fully rational creatures. Animals and madmen, on the other hand, do not even have this potentiality (*EN* 1149b34-1150a1). Thus, Sauvé-Meyer is not entitled to say that "the

---

[36] Sauvé-Meyer, S. (1993), p. 23. Her translation.

reason Aristotle gives here for excluding non-human animals and the insane from virtue and vice of character *is also* the reason for him to exclude children from these categories of moral evaluation".[37] It is, I think, very significant that she goes on to quote the following passage of the *EE* in support of her claim:

In a human being both <reason and desire> are present (*enestin*) – that is, at a certain age, the one at which we attribute action (*praxis*). For we do not say that a child acts, nor that a beast does, but only someone who is already such as to act due to reasoning. *EE* 1224a27-30

This passage, I think, must be interpreted in the light of Aristotle's claim, just a few lines after, according to which "*we possess by nature* both reason and desire" (1224b28-29). His argument to the effect that we possess (*echein*) reason by nature is that "of the two sources of action also reason is possessed[38] by nature, because if <human> development is permitted and not maimed, it will be present (*enestai*)" (*EE* 1224b29-30). Aristotle is here unambiguously claiming that children *have* reason (*echein logon*), albeit in a potential, or perhaps merely passive way (the latter being the emphasis in *EN* 1102b31-32).[39] By contrast, Aristotle seems to be using the verb '*eneinai*' ('to be present in') to refer to the actual possession of reason as exhibited by rational adults: only in them reason is *present*. This distinction between 'having' and 'being present' provides a clue to interpret *EE* 1224a27-30 supra. As suggested by Aristotle's use of '*enestin*' in 1224a27, his claim that in children reason is *not present* does not show that there is no sense in which they *have* reason (e.g. potentially). The mention of animals, along with children, is merely due to the fact that we do not attribute 'conduct' to animals either.[40] The fact remains that the reason why we do not say that that children and the reason why we do not say this of animals either, are not really the *same* reason: in children, reason is not present, because not actual or developed, whereas in animals, it is not present because it is

---

[37] *Ibid*. p. 23.
[38] See '*echomen*', 1224b29.
[39] As Aristotle says in *Pol.* 1260a14, the child *has* (*echei*) the deliberative part of the soul, but in an undeveloped form (*atelês*).
[40] Not even the *EE* claim that that we do not attribute conduct (*praxis*) to children is as conclusive as it may seem. The *EN* and the *MM* for instance only deny conduct to animals (*EN* 1139a20; *MM* 1187b7-9).

extraneous to their nature. In fact, the *EE* account of what it means to have a rational *nature* suggests that children, on Aristotle's view, are *not* 'outside rational nature' at all.

That children are not, in Aristotle's view, outside rational nature in the way non-rational animals or madmen are, is of the greatest significance for our present purposes, because for Strawson, children, psychopaths, and perhaps also beasts, are all non-moral agents *for the same reason*: it is as inappropriate to hold them accountable for their voluntary behaviour through holding the S-reactive attitudes towards them, as it is to deem them capable of holding others accountable. More importantly, agent-based considerations of the sort 'he is just a child', 'he is a psychopath', etc. are all, *for the same reason*, considerations that invite us to regard these agents in terms of an 'objective' attitude. This is why Strawson often merges together considerations like 'he is only a child' along with 'he is a hopeless schizophrenic', 'his mind had been systematically perverted', 'he is warped or deranged' or 'he is neurotic'.[41]

Now, in parallel with, and connected to, the Strawsonian blending of these importantly different agent-based considerations, the Strawsonian tradition ignores a fundamental distinction between *moral education*, teaching, and the sort of attitudes towards pupils involved in the formation of their character, on the one hand, and behavioural *control* and *conditioning*, which applies to psychopaths, animals and children alike. On the Strawsonian view, both education and conditioning involve instrumentalist (i.e. 'objective') attitudes. Sauvé-Meyer mirrors this view precisely on this point, when she says that, for all we know, "the praise and blame for which Aristotle thinks voluntariness is necessary might simply be *tools of behavioural control* and *character formation justified by purely prospective considerations*",[42] thus merging two importantly different attitudes, namely, plain behavioural manipulation, therapy, and conditioning on the one hand, and moral education ('character formation') on the other. As I have already suggested, the phrase 'prospective attitude' or 'prospective consideration' fails to capture the sense in which Strawsonian 'objective' attitudes are non-moral, because not all prospective attitudes are *ipso facto* 'objective' or better, instrumentalist. Roberts' prospective

---

[41] These are all Strawson's own phrases. See Strawson, P. F (1962), p. 78-79.
[42] Sauvé-Meyer, S. (1993), p. 40. My emphasis.

conditioning interpretation makes a parallel mistake, in the course of arguing that praise and blame as involving prospective conditioning can ground an account of Aristotle's theory of voluntariness. Roberts suggests that "[o]ne 'deserves' punishment, for Aristotle, if there is something wrong with one's soul of the sort *that might be correctable*. Thus *animals*, *small children*, and *fully mature adults* will all be in need of punishment *for the same reason*."[43] 'Correctable' here is the operative word, and it covers several importantly different things. It may cover the sort of praise and blame, and the sort of punishment and reward *purely* justified in terms of social utility, or at least purely understood as 'training' or 'conditioning'. If so, then of course animals, and perhaps even some psychopaths, in so far as they are responsive to pain and pleasure, would be included in the list of their 'appropriate targets'. But 'correctable' could also mean 'subject to moral training'.

That moral education is significantly different from the ill-regarded practices of conditioning, control and manipulation, is suggested by our refusing to identify it with these practices (or call it by the corresponding names). The most significant distinction between teaching someone and conditioning or controlling his behaviour, for our present purposes, is that teaching is for the good of the learner, whereas conditioning him is for some good, *not his own*.[44] At the limit, conditioning and controlling S's behaviour involves no consideration for S's actual or potential autonomy: it is, in this sense, purely instrumental. *Moral formation*, on the other hand, is for S's own sake, and is thus quite different from conditioning: it is not 'instrumentalist' in the sense in which treating someone instrumentally seems to us so at odds with the concerns of morality. Recall that plain behavioural manipulation, like the one used to train animals and to pacify psychopaths, is seen by Strawson himself as a way of excluding the targeted agent from one's moral community, or as a way of failing to treat the targeted agent as a moral individual in his own right. If *this* is the reason why we want to withdraw the adjective 'moral' from this sort of attitudes (i.e. the instrumentality argument) namely, because it

---

[43] Roberts, J. (1989), p. 23. Italics are mine.

[44] I take it that verbs like 'conditioning', 'controlling', 'manipulating' have taken on the pejorative meaning of 'regulating someone's behaviour in the interests of the controller, conditioner, etc. and not of the subject'. In itself, 'training' or 'conditioning' *could* include character formation. Their pejorative meaning is clearly associated with the Pavlovian idea of 'conditioned reflex'.

treats its targets in a merely instrumental way (as means to ends from which external agents, rather than the targets themselves, benefit) then this reason does not apply to moral education. Therefore, the instrumentality argument is a very poor reason to think that the sort of attitudes involved in moral education, and in particular the sort of praise and blame bestowed upon the pupil in the process of character formation, are *not* moral attitudes. If this is so, then the fact that, for Aristotle, children are *not* 'outside rational nature', at least not in the way animals and psychopaths are, is of the greatest significance. If the instrumentality argument offers no cogent reason to withhold the adjective 'moral' from the attitudes involved in character-formation, then it offers no cogent reason to think that the dividing line for Aristotle between the moral agent and the non-moral agent, must be set between (i) the rational, mature prohairetic agent in possession of an ethical disposition,[45] and (ii) the immature agent who has not yet developed an ethical disposition; rather than between (a) what is *potentially* rational (both (i) and (ii)) and (b) what does not have such potentiality to begin with: for educational attitudes bear only upon the former, and not upon the latter.

I am not denying that educational attitudes can be described as 'forward looking' or 'prospective'. What I am denying is that they are forward looking or prospective *in the relevant sense* (i.e. 'instrumentalist') that makes us, according to Strawson, withdraw the adjective 'moral' from them, namely, in the sense in which the 'objective' attitude is justified by mere considerations of social utility, considerations which, if applied to a moral agent, are seen as 'offending his humanity'. A parent who scolds a child in moral terms for not having kept his promise may do it partly in order to deter him from failing to keep his promises in the future. But he is doing it for his child's *own good* or, in Aristotelian terms, for the sake of his own ultimate *aretê* and *eudaimonia*. He is helping him to *become* an autonomous agent. Thus, from the mere fact that an attitude is forward-looking or prospective it does not follow that it is at odds with the concerns of morality; for not all forward-looking attitudes are justified merely in terms of social utility.

---

[45] I am including continent and incontinent agents in this category, as Sauvé-Meyer does.

Now, I suspect that the Strawsonian interpreter may retort as follows: "I grant you that I have misrepresented the peculiar, non-instrumental nature of educational attitudes. But there is still a sense in which agent-based considerations of the sort 'he is just a child' belong to the same category as 'he is a psychopath' or 'he is deranged', to wit, these are agent-based considerations that make us *suspend* the S-reactive attitudes towards these agents in a sustained way". From this we are expected to conclude that children are *not* moral agents and our attitudes to them are *not* moral ones. But if we have indeed educational attitudes to children that can be characterized as 'moral', one can arrive at a quite different conclusion. Part of what is essential about educational attitudes is that the sort of blame exemplified by the father's scolding his child has to belong in some way to *the same category* as the sort of blame that the child *himself* is meant to bestow upon the same type of actions – and ultimately upon their agents – when these become of his concern once he has acquired full membership into the moral community. And by 'the same category' I mean here that educational attitudes are more than a mere *mimêsis* of the genuine reactive attitude emotionally toned (i.e. imitations of the overt behaviour, face, posture, voice, etc. involved in the 'original').[46] If the sort of blame exemplified by an educational attitude is to belong to the same category as the sort of blame that the child *himself* is meant to bestow upon others once he has acquired full membership into the moral community, it must be *in the same* category as the latter in a sense that is strong enough for the educational attitude to promote *education*, and not the mere repetition of patterns of external behaviour. It follows from this that the Strawsonian theory is very far from giving us the whole truth about our *moral* responses to human agents.

Moreover, it is not even clear that Strawsonian himself can *consistently* claim that immaturity is the sort of agent-based consideration that inhibits in a sustained way (as psychopathy does) the S-reactive attitudes, and to conclude from this that children are not moral agents and our attitudes to them are not moral ones. Strawson himself [47] pictures the treatment of children involved in their moral education and character

---

[46] For instance, in the way Aristotle thinks "most children's games should be *mimêseis* of the serious occupations of later life" (*Pol.* 1336a34).

[47] A similar inconsistency is evidenced in Wallace, R. J. (1994): compare his claims at p. 118, p. 114, p. 155 with his claim at p. 167.

formation as one that gradually, though 'insensibly', shifts between objectivity of attitude and genuine reactive attitude. Parents, he says: "are dealing with creatures who are potentially and increasingly capable both of holding, and being objects of, the full range of human and moral attitudes, but are not yet truly capable of either. The treatment of such creatures must therefore represent a kind of compromise, constantly shifting in one direction, between objectivity of attitude and developed human attitudes. Rehearsals insensibly modulate towards true performances."[48] For the purposes of the ongoing argument, we need not push the point further, for *even on Strawsonian grounds*, it is not at all clear that children and the attitudes involved in their moral education are simply not moral (because non-S-reactive).

At any rate, the Strawsonian argument is not that educational attitudes (and their targets) are not moral, because they belong to a different category from the distinctively moral S-reactive attitudes - a claim that I have argued against in the previous two paragraphs, anyway - but rather that that they are not moral because instrumentalist, and I have offered (I think) convincing arguments against this. If I am right, these arguments undermine one of the main motivations for trying to find in Aristotle an *exclusive* concern with those actions that originate from a constituted, mature ethical character and with a sort of praise and blame that is exclusively bestowed upon agents in possession of mature ethical dispositions and the actions issuing from them: since the sort of praise and blame that is involved in moral education is thought by the Strawsonian interpreter to be on a par with manipulation and behavioural control, it is felt necessary to establish first that Aristotle displays such an exclusive concern with mature ethical agents and their ethically significant actions, in order to show that his concern is a concern with genuine moral responsibility. Severing moral education and educational attitudes from sheer behavioural control, manipulation and therapy, and from the sort of 'objective', detached attitudes involved in them, undermines this assumption by Strawson's own standards. Moreover, and this is very important, it is now not so clear *what* this notion of a morally responsible agent really is. I will return to this point in Section E of Chapter 2.

---

[48] Strawson, P. F. (1962), p.88

In order to secure the *moral* status of Aristotelian praise and blame, by claiming that they are *retrospective* in nature, the Strawsonian interpreter, it seems to me, can still try to show that Aristotelian praise and blame are attitudes that *look back* (as the S-reactive attitudes do) to the agent as someone who 'deserves' or 'merits' such an attitude in virtue of an action of his that has breached or fulfilled a *moral demand* (e.g. Strawson's basic moral demand of goodwill).[49] This is, as we noted in Section B, the other alternative open to the Strawsonian interpreter. In the next section, I want to suggest that this alternative is also closed.

D. The morality of character traits

An initial argument for this negative conclusion can be derived, I believe, from Aristotle's notion of ethical disposition or character trait (*êthos*). The use of the adjective 'moral' carries strong implications that would not have been endorsed by Plato and Aristotle. Aristotle (and Plato) did not distinguish between what many ethicists would call 'moral virtues' (that is, other-regarding or cooperative excellences) and self-regarding ones.[50] Moreover, had such a distinction been drawn by the time Aristotle was writing the *Ethics* he would probably have considered it irrelevant for the strict purposes of defining *êthê* or ethical *hexeis* (ethical dispositions). Of course, Aristotle has much to say with regard to other-regarding excellences of character like justice, the excellences associated with friendship, or virtues like courage, generosity, open-handedness, or truthfulness, and some of their corresponding vices, all dispositions that are 'other-regarding' at least in the narrow sense that they typically bring benefits to other people or produce direct harm to them. Still, what is important about excellence of character (*êthikê aretê*), in Aristotle's

---

[49] Recall Sauvé-Meyer's claim that reactive attitudes "are *based on certain expectations of interpersonal regard* and they express our reactions to the fulfillment or non-fulfillment of those expectations" and that "to be subject to these expectations is to be subject to the *demands of morality*". Suavé-Meyer, S. (1993), p. 18.

[50] So for instance, Brandt (1970) registers a use of 'moral character' that philosophers have had in mind when they have accused Aristotle of discussing, in his account of the virtues, some traits that are not traits of *moral* character, like ready wit. One important feature of traits of character in the sense these critics have in mind is, according to Brandt, that each of them is "either an important asset or an important liability *for cooperative living*, from *the point of view of society*" (p. 24, my emphasis). See also Adkins, A. W. H. (1960), pp. 336-340 for an account of how the cooperative virtues came to be included in the schema of *aretai*, without taking precedence over the non-cooperative ones.

view, is a feature of it that connects it primarily with its possessor: it is in the possession and, more conspicuously, exercise of excellence of character that human beings attain the perfection appropriate to their own condition *qua* human beings.[51]

There are many examples of what I call 'self-regarding' ethical *hexeis*, i.e. ethical dispositions where assessment of the corresponding conduct focuses on it in relation to the agent himself as distinct from foreseeable effects on others. Perhaps ambition, moderation or greatness of soul are good examples on the side of virtue. Wastefulness or self-indulgence are, I think, good examples on the side of vice, and Aristotle himself seems to be well aware of this self-regarding aspect in the case of vices corresponding to the virtue of munificence, like being *banausos* by spending too much money on irrelevant occasions, or being shabby: he claims that, although these are faulty dispositions (*kakiai*),

they do not actually attract people's reproaches (*oneidê*), because they are neither harmful (*blaberai*) to one's neighbor nor excessively discreditable. *EN* 1123a32-33

The same can be said with regard to the vices corresponding to greatness of soul, that is, being little-souled and being conceited: Aristotle observes that

these people too are thought of not as bad (*kakoi*), for they do not harm anyone (*ou kakopoioi eisin*), but as having got things wrong (*hemartêmenoi*). *EN* 1125a17-19

By explicitly saying that these dispositions do not bring social reproach or censure (*ouk oneidê epipherousi*), and that they are not harmful (*blaberai, kakopoioi*), Aristotle declares his conviction that, although current social practices of praising and blaming should not perhaps be totally ignored, they should not be taken as a criterion for determining which dispositions are excellent and which ones faulty. This conviction of Aristotle's is relevant to our argument in two connected ways. The first one is this: by

---

[51] A good example of how the virtues or vices would be ranked or classified from the point of view of their social benefit and harm (i.e. from the point of view of popular morality), according to the criterion of other-regard, is Aristotle's recognition of the 'popular' view of virtue in the *Ars Rhetorica* as "a power of providing and preserving good things, a power productive of many and great benefits to all men in all cases" (1366a4-b5).

'moral demands' (i.e. moral obligations) Strawson and the Strawsonian theorist understand, *inter alia*, the sort of demands the breaching of, or the acquiescence in which makes it appropriate or justifies the corresponding reactive attitudes. But now ask yourself whether it would be appropriate to feel moral resentment or indignation, or to believe that such attitudes would be appropriate, towards a healthy adult who spends too much money on irrelevant occasions, or towards a glutton. Well, not necessarily. But this just means that the sort of blame appropriate to these agents on the occasion of their faulty behaviour is not, according to Aristotle, made 'appropriate' by the fact that it breaches a moral demand, in the Strawsonian sense of 'moral demand'. This is precisely what Aristotle implies when he declares his belief that, although current social practices of praising and blaming should not perhaps be totally ignored, they should not be taken as a criterion for determining which dispositions are excellent and which ones faulty, or which ones are to be *praised* and which ones *blamed*, in *his* sense of praise and blame.

Another way to express the same point (and limiting oneself to the case of vices, as Aristotle does), is by emphasising the fact that self-regarding vices are not manifested in conduct that (qua expressing such dispositions) *affects others*, at least not in the sense in which its affecting others would make it appropriate for them to hold the S-reactive attitudes. Hence, such conduct is not 'blameworthy' in the Strawsonian sense that its agent merits or deserves the relevant sort of S-reactive response. Again, this is what Aristotle implies by saying that self-regarding vices do not bring social reproach or censure, and that they are not harmful.

Now, if the vices corresponding to virtues that are self-regarding and to the ones that are other-regarding are all blameworthy *for the same reason*, then it follows that Aristotelian praise and blame are *not* attitudes that look back (as the reactive attitudes do) to the faulty (or excellent) agent as someone who deserves or merits such an attitude in virtue of an action of his that has breached (or fulfilled) a *moral demand*, in the Strawsonian sense of 'moral demand'.

The second way in which Aristotle's conviction is relevant to our present argument is this: it follows from the aforesaid conviction that Aristotelian praise and blame do not

correspond to the reactive attitudes, for if (i) existing social responses of reproach, as reported by Aristotle, are typically aimed at those faulty traits of character that are directly harmful to other people; and (ii) it is reasonable then to think that they are typically the manifestation of S-reactive attitudes like resentment or moral indignation; (iii) and since Aristotle believes that an ethical disposition can be considered as a vice and as something blameworthy regardless of whether it attracts these sort of reactive responses or not, then (iv) it is reasonable to think that the sort of blame associated with vice is not identical with a S-reactive attitude like resentment or indignation.

As a result, it is fair to say (granting the argument in Section C) that Aristotelian praise and blame cannot be shown to be *moral* (in a sense that does not translate *'ethikos'*) *via* establishing its retrospective nature, as the Strawsonian intends: for 'prospective' is not *eo ipso* non-moral, and Aristotelian praise and blame do not 'look back' to the agent as someone who deserves or merits them, in virtue of a piece of behavior that has breached or complied with a (non-tautological) *moral* demand. Moreover, the second feature of the Strawsonian interpretation (i.e. the concern with the moral agent) is, once more, inevitably thrown into doubt. For if the praise and blame Aristotle has in mind do not correspond to the S-reactive attitudes, then it is not at all clear what is the motivation for trying to find in Aristotle an equivalent of the Strawsonian moral, answerable agent, i.e. the agent to whom such attitudes are appropriately addressed, for it is now clear that the Aristotelian virtuous or faulty agent is not *this* moral agent (even if these two concepts have the same extension).

As we shall see in the next section, Aristotle wants to say that praise (*epainos*) is bestowed upon virtuous people, and blame (*psogos*) bestowed upon faulty people, and to keep virtue and praise, as well as vice and blame, as connected as possible (*Rhet.* 1367b27-33; *EN* 1101b33-5; *EE* 1219b8-17). If the praise and blame in question were to correspond neatly to social esteem and condemnation, Aristotle would be contradicting himself, in stating that social esteem and condemnation are not an adequate criterion for the praiseworthiness and blameworthiness of virtue and vice. If, on the other hand, not all praise and blame is manifested in the form of honours (*timai*), the corresponding sense of

worth and pride, and social reproaches (*oneidê*) and the corresponding sense of guilt, no such contradiction would ensue. Praise and blame are primarily manifested, as I will show, in the judgement or opinion (*doxa*) of a spectator and, most importantly, in language itself. Recognizing these levels of praising and blaming is important, because it allows Aristotle to say that, in a sense, to be ambitious in the right manner or to be pusillanimous are 'praiseworthy' and 'blameworthy' things respectively, without implying that they are (or should be) dispositions that are open to typical S-reactive attitudes like reactive social esteem or condemnation.

E. Aristotelian praise and blame

The next group of arguments against the Strawsonian interpretation, in Section F, is based on a more or less detailed account of Aristotle's own conception of praise and blame. Providing this account is the task of the present section.

Plato in the *Republic* and the *Laws* already had employed a very technical notion of praise (*egkômion*, *epainos*) and blame (*psogos*). The kind of poetry that Plato would have admitted in his Republic was confined to the production of "hymns to the gods and praises of good men (*enkômia tois agathois*)" (*Rep*. 607a). The goodness in this case is not confined to the moral domain. In the *Laws*, in the process of describing the organization of the festival-contests, Plato says that people should "compose for one another speeches of praise and blame (*enkômia te kai psogous poiein*), according to the character each one exhibits not only *in the contests*, but in his life generally, magnifying him who is accounted most good and blaming him who is not" (829c), and he makes clear that these are lyrical speeches or songs.

If there is an Aristotelian canonical account of praise and blame at all, this is the one that distinguishes praise (*epainos*) from encomium (*egkômion*). This distinction (which Plato, as far as I know, did not recognize) is ubiquitous in Aristotle's writings. The *Ars Rhetorica* explains it thus: "Praise is the recital that exhibits the greatness of excellence (*logos emphanizôn megethos aretês*), and therefore we must display a man's actions as

excellent. Encomium refers to actual deeds…" (1367b27-30). The distinction is confirmed by the Eudemian claim that "praises are appropriate to excellence (*epainoi tês aretês*) through deeds (*dia ta erga*), and encomia appropriate to deeds" (1219b8-9), and its definition of these notions: "an encomium is a recital (*logos*) of a particular deed, praise is a recital of the general quality of a man (*tou katholou*)" (1219b14-16). It is further verified by the *EN*: "Praise is appropriate to excellence, for as a result of excellence (*apo tautês*) men are disposed to do fine actions (*praktikoi tôn kalôn*); but encomia are appropriate to deeds (*erga*), whether of the body or of the soul" (*EN* 1101b31-4).

For our present purposes, the most important characteristics of *epainos* highlighted in these passages are that: (A) *epainos* is a form of *logos* (a speech act, one could say) which (B) aims at exhibiting the eminence or magnificence of someone's virtue as a general disposition (i.e. as opposed to particular virtuous actions). (C) There is another feature of praise and blame (implicit in the Platonic background) that is not explicit in the passages just quoted, but that comes to the fore when the *EE* illustrates the claim that praises are appropriate to excellences through deeds, by pointing at the fact that "it is the victorious who are crowned (*stephanountai hoi nikôntes*), not those who are capable of winning but do not win" (*EE* 1219b9-10), and the *EN* claim that "we praise the just or brave man and in general both the good man and excellence itself because of the actions and functions involved, *and we praise the strong man, the good runner, and so on*, because he is of a certain kind and is related in a certain way to something good and important" (*EN* 1101b12-18). What these passages show is that the praise and blame Aristotle has in mind is not peculiar to the domain of the ethical, but extends also to the domain of the technical.

The first two features of praise (and, let us suppose, also blame) just mentioned are somehow cryptic. Let me start with (B), the claim that praise aims at exhibiting the eminence of virtue *as a disposition*. Typically, Aristotle employs two expressions: (i) '*ho epainos tês aretês*', or (ii) '*aretai epainetai*'. (Aristotle seems to use no expression parallel to (i) in the case of vices, but he does say that vices are blameworthy, e.g. '*kakia* is among the *psekta*', *EE* 1223a10). In saying that ethical dispositions, virtues and vices, are 'worthy'

of praise and blame, and presenting virtue in particular as the ultimate target of praise,[52] Aristotle is clearly referring to virtues and vices *as qualities*.[53] By this I don't mean that the primary object of praise and blame is Virtue and Vice personified, as it were. I am rather making the logical point that the immediate, or primary, object of praise and blame is the virtue or the vice rather than the person who has them (i.e. the substance qualified by them); that is, Aristotle is saying that *virtue* is praiseworthy and *vice* is blameworthy, not that virtuous and faulty individuals like Socrates or Morris Townsend are praise- or blameworthy respectively. This latter may be also the case, and even follow from the first claim, but it is not what the claims "praise/blame is of virtue/vice" or "virtue/vice is praiseworthy/blameworthy" mean.

Secondly, by these expressions, (i) and (ii), Aristotle may give the impression that *first* one identifies something as a virtue or a vice, and *then* one decides to praise or blame it (or its possessor), which is clearly absurd, given that virtue- and vice-*terms* ('courageous', 'unjust', etc.) *themselves* carry a laudatory and pejorative *meaning* respectively. Calling someone 'courageous' is *already* to praise him, and to call someone 'coward' is *already* to blame him (i.e. we do not first identify him as courageous or coward and then praise or blame him!). If virtue- or vice-terms refer primarily to *qualities*, this fact about virtue- and vice-terms should not be surprising, given that these qualities are highly regarded or depreciated respectively. So I take it that the expressions '*ho epainos tês aretês*', or '*aretai epainetai*' have two aspects, one being that *qualities* like virtue or vice have value or disvalue respectively (and transfer this value or disvalue to their possessors), and the other being that virtue- and vice-*terms* (perhaps because they refer to these qualities) have laudatory and pejorative meanings respectively. Let me say something more about this semantic aspect of the claim that 'praise/blame is of virtue/vice' (the point may seem obvious, but it has never, to my knowledge, been explicitly made by scholars).

---

[52] *EN* 1101b31-2, 1103a10, 1108a14, 1126b65-6; *EE* 1219b8, 1220a5-6, 1223a10-13, 1228b30-1, 1233a4-8; *Rhet*. 1366a37.

[53] Virtues and vices are *hexeis* (habits or stable dispositions), and *hexeis* are in the category of quality. See *Cat*. 8b25ff.

Aristotle clearly thought that praise and blame were common-sense *tests* for considering a given trait of character as an excellence or a defect. So for instance at *EN* 1103a4-10, when Aristotle is arguing for the distinction between intellectual and ethical dispositions, he says with regard to wisdom (*sophia*, an intellectual virtue) that it is not an ethical excellence, but still it is an excellence: "*because* we praise the wise person because of his disposition, *and those dispositions which are worthy of praise we call excellences*". The *MM* also registers a similar argument aimed at showing that practical wisdom is an excellence: "practical wisdom is an excellence … *because* wise people are praised, and praise is appropriate to excellence (*epainos aretês*)" (1197a17).[54] If this is partly what Aristotle means by the possibly misleading expression 'praise is of excellence', or by saying that virtue and vice are praiseworthy and blameworthy respectively, then this is in fact a strong reason for favouring an interpretation of 'praise is of virtue' and 'blame is of vice' as 'virtue-terms and vice-terms have laudatory and pejorative meanings respectively'. We have seen that Aristotle cannot mean by this something like social esteem and reproach, at least as understood as expressions of S-reactive attitudes (Section D). What better employment could we then make of the idea that 'praise' and 'blame' are used as a test of excellence and vice than to ask ourselves whether the name of a candidate for excellence or vice carries or not a laudatory or a pejorative sense respectively? The same test is used by Aristotle in *EN* II 4 with the purpose of showing that excellences and bad states in general are not affections (*pathê*). Aristotle says that we are *called* (*legometha*) 'excellent' or 'faulty' after (*kata*) the virtues and vices (*EN* 1105b29-31), and not after mere affections. He makes it clear that by this he means: "we are *praised* (*epainoumetha*) or *blamed* (*psegometha*) after (*kata*) the virtues and vices" (1105b32-1106a2). Again, it is clear that he means by this that mere affections-terms like 'angry' or 'confident' do not convey a laudatory or pejorative *meaning*.

Finally, let us say something about the first feature, (A), of Aristotelian praise and blame, namely, their being characterized as '*logoi*'. More specifically, Aristotle understands praise as a *logos* exhibiting a man's excellence (a claim advanced both by the

---

[54] Also *EN* 1106a6-10.

EE and by the *Rhet*.),[55] and perhaps also blame as a *logos* exhibiting a man's defect of character. The closest Aristotle gets to explaining the claim that praise and blame are *logoi* is in the *Rhet*. 1367b36-1368a9, when comparing praise with counsel or advice (*hai sumboulai*):

Praise and counsels share something in common (*echei koinon eidos*): for what you might suggest in counselling becomes praise when differently expressed. If therefore we know what things one ought to do and what sort of character one ought to have, we must make an adjustment in the phrase (*tê*[i] *lexei*) and reverse it by expressing these things as suggestion (*ôs hupothêkas*). So for instance <the statement> that 'A man should be proud not of what he owes to fortune but of what he owes to himself', when expressed in this way has the force of a suggestion (*hupothêkèn dunatai*); but expressed thus 'He is proud not of what he owes to fortune, but of what he owes to himself', it becomes praise. Accordingly, if you want to praise, look at what you would suggest, and when you want to suggest, look at what you would praise. The expression (*lexis*) will necessarily be of one or other of two opposite forms [i.e. a virtue or a vice], depending on whether (*hotan*) the preventive form (*to kôluon*) or the non-preventive form (*to mê kôluon*) <of the suggestion> is adjusted. [56]

Aristotle is clearly not saying that the *meaning* of 'he is proud of what he owes to himself' is captured by the fact that the sentence is used to praise someone. *If* you want to praise, then you can use the sentence 'he is proud of …' for such purposes. The sentence 'he is proud of what he owes to himself' (which has the same referent as 'he is magnanimous') standardly used for such purposes as evaluating or praising someone, cannot be said to have such meaning (an example is when the same sentence occurs in a subordinate clause such as 'If he is magnanimous, then we should kill him'). Praise and blame in this sense are what speech-act theorists would call 'illocutionary forces' (as distinct from locutionary

---

[55] Here Aristotle follows an old tradition. Besides Plato's *Laws* 829c see also *Menexenus* 234c-235c, where Socrates describes the praise given to someone who has fallen in battle, in the form of a speech (*logos*) prepared long beforehand. Also interesting about the *Menexenus* is that it clearly shows how one of the essential functions of praise is the appeal to children to copy the virtues of the heroes (236e, 246bff).

[56] This passage presents some problems of interpretation. It is not clear whether the final reference to *to mê kôluon* (i.e. allowance?) can be translated as a positive command to do something (as it is suggested by the examples Aristotle gives), nor is it clear whether Aristotle is in fact distinguishing between a counsel or suggestion (the proper translations of '*sumboulia*' and '*hupothêkê*') on the one hand, and a *command* (this latter suggested by the use of '*dei*' and the final reference to *to kôluon*) on the other. It is not necessary, however, for us to take a decision with regard to these issues.

forces or meanings). They occur as/in speech acts and not as/in the meaning of words apart from the use we make of them. Aristotle seems to be conscious of this, for he distinguishes '*the things that* you might suggest *in* counseling' ('suggestions') from the grammatical form called 'counsel' (see also *hupothêkèn dunatai*, 1367b36).[57] Aristotle's point in the passage is that praises and suggestions are different illocutionary acts expressed through different moods, having in common their propositional content: the propositional content of what is praised, the *eidos* (e.g. being proud only of good things owing to oneself) is identical to what is suggested, and the difference lies in the moods, the first one in the indicative, and the second one in the imperative or gerundive mood. One can be used as a criterion for the other and *vice versa*.

F. On why Aristotelian praise and blame tells against the Strawsonian interpretation

This, then, is Aristotle's account of the nature of praise and blame: they are speech acts implemented through sentences in the indicative mood, and containing V-terms (the 'atomic units' of laudatory meaning) or descriptions grounding V-terms predicated of one or more individuals, on the occasion of his or their recognizably v-deeds. Notice that Aristotle seems to think that praising and blaming *are* preeminently judgements *in words* (i.e. *lexeis*) of people's or actions' praise- or blameworthiness and that expressing these judgements is the same as actually praising or blaming people for their actions. This is clear from his comparing them with counsels or advices, which are obviously addressed in words to others. This is an important feature of Aristotelian praise and blame, and one that we shall have occasion to discuss in Chapter 2.

One difference between Aristotle's notions of praise and blame and S-reactive attitudes immediately emerges from this understanding of praise and blame as *logoi*. Aristotelian praise and blame are just speech acts. Nowhere in the Aristotelian account of praise and blame do we find the slightest trace of the S-reactive attitudes and the sort of emotional responses so important for the Strawsonian model. This is all the more

---

[57] This may be obscured by the fact that Aristotle does not use two different expressions to refer both to the speech act of praise and to the praise-sentence (the praise-*lexis*).

remarkable given the complexity of Aristotle's own view on the emotions and the richness of vocabulary at his disposal to describe emotional phenomena, as shown by his ethical treatises, the *Ars Rhetorica* and the *Poetica.* For example, the Greek attitude that seems to be the closest to Strawsonian moral resentment is *orgê*: a painful desire for revenge or punishment, caused by a slight (which is by definition intentional or voluntary) to oneself or to one's relatives or friends (*Rhet*. 1378a30ff). It is also said by Aristotle to be prompted by "slights from those by whom one thinks one must (*dein*) be well treated" (1379a7-8), and to be checked by the recognition that the supposed slight was involuntary (1380b33). Nowhere in the *corpus* is *orgê* associated with blame.[58] Furthermore, *orgê* is clearly not suitable for performing the functions attributed to moral resentment, since it only seems to operate within one's narrow circle of acquaintances, rather than the whole of one's moral community.

Now, one may think that this particular stress on certain intrinsically moral emotions or the belief in their adequacy is a feature peculiar to Strawson's own conception of moral responsibility, and accordingly, that the argument on the previous paragraph, and some important aspects of the arguments offered on the previous sections, are far too provincial for them to be of great significance. So let us single out those features of the Strawsonian account that display it as a particular version of a wider conception of moral responsibility, that is, as *accountability*. As we have seen, moral responsibility, on the Strawsonian view, is defined by the practice of *holding people responsible* (i.e. accountable); this is because being morally responsible for an action is, on this view, being an appropriate target of interpersonal S-reactive attitudes, and these are emotions that we would feel, or believe it would be appropriate to feel, when moral expectations are breached. Because of this, interpersonal S-reactive attitudes are *ways of*

---

[58] Another possible candidate for performing the functions that moral indignation performs in the Strawsonian model is *nemesis* ('righteous indignation'), portrayed by Aristotle as a virtue. *Nemesis* is the pain at undeserved good fortune (*EN* 1108b3-4), and lack of pain or rejoice at deserved bad fortune. Not only is *nemesis* never associated with blame; it is also clear that it cannot be thus associated because it is a response to the things that *happen* to our neighbors according to their worth – not to the things that our neighbors *do*. If anything, it is a response to seeing people being rightly sanctioned or punished, not itself a form of punishment or sanction (on *nemesis* see *Rhet*. II 9).

holding moral agents responsible, *accountable* or *answerable*, for breaching (or failing to meet) these moral expectations.[59]

The first point that emerges from stressing this aspect of the Strawsonian conception of moral responsibility as accountability is the particular emphasis that it places on *negative* S-reactive attitudes: it turns out that S-reactive attitudes are primarily those emotions that we feel or believe it would be appropriate to feel towards someone whose actions have *breached* our moral expectations (i.e. what we have a right to expect from others). This emphasis on the negative S-reactive attitudes naturally suggests some kind of asymmetry between moral responsibility for praiseworthy actions and the one for blameworthy actions. The reason for this asymmetry is that, as Wallace points out, "in the case of responsibility for worthy acts, we do not suppose there to be any particular positive sentiment that we are or ought to be subject to".[60] So, even if the accountability theorist can somehow accommodate the fact that people are also *responsible* for worthy deeds,[61] i.e. for those deeds that fulfill (or even exceed) our moral expectations, whether or not people are also positively *praiseworthy* for these deeds and *credited* for them, seems rather irrelevant from the accountability point of view.

Consider the sort of legal accountability we are so familiar with in modern liberal societies. If you breach the law, you are of course liable to legal sanctions – you are held answerable for breaching it. If you fulfill your legal obligations, on the other hand, you are responsible for doing so in the sense in which you are an *accountable subject* (someone who can, *in general*, be held legally answerable for his illegal actions) and your fulfilling your legal obligations is deemed voluntary (questions about the coercive aspect of the law

---

[59] It should be clear now that, on the Strawsonian view, 'moral agent', 'moral attitude' and 'moral obligation' are defined in terms of each other. This is more explicit in R. Jay Wallace's sophisticated Strawsonian account of moral responsibility.

[60] Wallace, J. (1994), p. 71. This tendency to focus on negative attitudes can be detected in Strawson's article (i.e. in his focus on resentment and moral indignation), and I think it is one of the recognizable features of accountability. See also Wolf, S. (1990). Ryle, G. (1949) seems to have been the first one to have paid serious attention to it.

[61] Wallace's account of responsibility for worthy acts is based on a dispositional use of the notion of holding accountable. On this account, what we do in holding A responsible for worthy actions is: (i) to regard A as an accountable agent (i.e. to be *susceptible* to reactive emotions or to believe these to be appropriate in the case A breaches the obligation); and (ii) to believe A has done something that meets (or exceeds) the shared moral obligations (*ibid*. pp. 70-71).

aside). But what you are doing is fulfilling your legal obligations, i.e. what is expected from you: even if you are not, accordingly, blameworthy, it is hard to make sense of your being deemed positively *praiseworthy* for doing so. After all, the state does not positively reward us for fulfilling our civic or legal duties. Now, if this is true of accountability, then it just means that accountability seats badly with Aristotle's particular concern with praiseworthiness (or at least with the fact that, for him, praiseworthiness and blameworthiness are plainly symmetrical). Clearly, the praiseworthiness of virtue is partly grounded on the fact that virtue is something great or eminent (*megethos*); that becoming virtuous and excel oneself is peculiarly hard to achieve and an accomplishment (*EN* 1137a9-17), and in particular, that virtuous activity normally goes *beyond* what is to be expected (*para to prosêkon*, *Rhet*. 1367b14-15). If the praiseworthiness (i.e. the value) of virtuous qualities is partly attached to these facts, then this is very significant, for fulfilling one's obligations and doing what is merely *expected* (*prosêkon*) from one as an accountable agent is not an achievement in this sense. Fulfilling moral expectations is a basis to be deemed a good agent, perhaps respected and trustworthy, but hardly praiseworthy in the sense Aristotle's virtuous agent and his deeds are.

Another central aspect of the Strawsonian paradigm in terms of accountability is even more significant, namely, that the negative S-reactive attitudes (which are primary or more basic) are naturally seen as dispositions expressed in *sanctioning responses* not merely restricted to moral blame, but also extended to "avoidance, reproach, scolding, denunciation, remonstration, and (at the limit) punishment",[62] that is, as involving dispositions to treat others in "generally unwelcome ways."[63] Furthermore, given this network of mutual demands between participants of a moral community and the corresponding view of people's responses towards conduct that breaches these demands as *sanctions*, the Strawsonian view naturally generates an interpretation of it in terms of fairness and desert: when is it *fair* (i.e. 'appropriate') to react unfavourably to someone's

---

[62] Wallace, J. (1994), p. 54.
[63] Watson (1996), p. 238.

behaviour?[64] When does an agent *deserve* an adverse treatment in response to his faulty conduct?

That Aristotle never refers to the notion of desert or merit when connecting praise and blame to voluntary actions and ethical agents, is thus extremely significant.[65] Suppose Aristotle were to say that S *deserves* blame for S's voluntary blameworthy actions: how would he express such claim? He would have said that one is *axios* of blame[66] in virtue of one's voluntary actions, or that blame is bestowed upon one *kat' axian* (in accordance with what one deserves), but he never says such a thing.[67] This suggests that his account of what it means to blame S *dia x* (*EE* 1219b8-9, *EN* 1101b14-16) or *epi x* (*EN* 1109b30-1, 1110a23, 1127b18) is altogether different from the accountability account, that is, '*dia x*' or '*epi x*' does not characterize that piece of behaviour *x* that makes S *deserving* of blame.[68] What these expressions mean is something I will discuss in Chapter 2; the important point here is that considerations of fairness and desert are for Aristotle unimportant in this connection, and this is pretty strong evidence that he is not concerned

---

[64] On this see Wallace, R. J. (1994), Ch. 4.2 - 4.3; and Watson, G. (1996).

[65] To indicate at this point that Aristotle apparently connects blame with chastisement (*kolasis*) at the opening of *EN* III (1109b31-35), does not by itself show that blame (*psogos*), on Aristotle's view, belongs to the same family of punitive attitudes as *our* notion of punishment or legal sanction. If anything, Aristotle's notion of *kolasis* as a cure effected by the infliction of *pains and pleasures* (*EN* 1104b13-18), and the connection between *kolasis* and childishness (*EN* 1119a33b15; 1180a2-4) tends to support Roberts' prospective conditioning interpretation. But even so, it must be first shown that there is indeed an intimate connection between praise/blame and chastisement/reward at 1109b31-35, rather than a mere hint at a further application of a theory of voluntariness, given that voluntariness is *also* a criterion for applying chastisements and rewards. The *EE*, for instance, does not refer to the institutional practice of chastisement and rewards in this context.

[66] This (accountability) use of '*axios*' is well documented. The defence in Antiphon's *First Tetralogy* (II. 11) argues that it is the prosecutors "who *deserve* (*axion*) in full the penalty"; Gorgias in his *Encomium of Helen* (6) says that "the barbarian who undertook a barbaric undertaking in speech and in law and in deed *deserves* (*axios*) to receive accusation in speech (l*ogô*[*i*] *aitias*), debarment in law (*nomô*[*i*] *atimias*), and punishment in deed (*ergô*[*i*] *zêmias*)"; and Demostenes (*Against Meidias* 42) observes that "the laws treat the voluntary and insolent transgressors (*tous hekousiôs kai di' hubrin plêmmelountas*) as *deserving* more resentment and a heavier punishment (*meizonos orgês kai zêmias axiousi*) than other classes of offenders." Note the (accountability) connection between *desert* and *punishment*.

[67] Aristotle thinks that the things distributed according to worth (*axia*) are external goods (*EN* 1123b17; *Rhet.* II 1387a8-16). Examples are honour (*EN* IV 3), love (*EN* 1158b27), political power (*EN* 1160b33), high birth and wealth (*EE* 1249a7-10). He makes it clear that 'worth' (*axia*) here means *positive* worth (*EN* 1123b17), and this rules out blame by definition. Secondly, even though there are affinities between praise and honour – so that it is not obvious why praise is not an external good, since honours are – Aristotle does not seem to think of praise as an external good.

[68] E.g. Sauvé-Meyer talks about "[t]he sort of praise and blame the agent *deserves for* (*epi* + dative) these actions and feelings", Sauvé-Meyer (1993), p. 51. Italics are mine.

with accountability: in particular, it clearly shows that for him blame does not express sanctioning dispositions. Note that I do not mean to imply that it does not matter to Aristotle whether blame or praise are *appropriate* to a given target (sometimes 'desert' is used in this general fashion); the point is that appropriateness does not necessarily consist in being deserved, in the particular sense of 'just deserts' (which comes under particular justice). Getting what you deserve in this narrow sense is just one of many forms of being treated 'appropriately'.

The last feature of Aristotelian praise and blame noted in the previous section is perhaps the most significant of all. It is clear that the Aristotelian notion of praise and blame applies to skills and their lack as well as to excellences and defects of character, and that it does this indiscriminately. This much we could already tell from Aristotle's mention of the victorious in a contest at *EE* 1219b8-17 and the strong man and the good runner at *EN* 1101b12-18 in the context of his account of praise (cf. C in p. 47). Such an indiscriminate application of praise to skills and ethical dispositions should come as no surprise, given the Platonic background sketched in Section E. It strongly suggests that when Aristotle claims that excellences and defects (*kakiai*) are the main (formal) 'objects' of praise and blame he does not seem to be referring exclusively to ethical dispositions, nor even exclusively to ethical and intellectual dispositions like practical wisdom, but also to skills (*technai*). After all, given the wide range of applications of the word '*aretē*' [69] it was natural for Aristotle to align ethical excellences together with intellectual excellences and skills in this regard.

Now, this feature of Aristotelian praise and blame is important, because the good runner's ability to win a difficult race or the bad musician's inability to play in tune make them praiseworthy and blameworthy respectively in a sense which significantly differs from the sort of praiseworthiness and blameworthiness appropriate to agents whose actions we believe have fulfilled or breached moral expectations, and for whom certain forms of moral attitudes like, let us suppose, admiration or resentment are appropriate. Of course, judging a musician bad could involve certain previous expectations or demands,

---

[69] Cf. *EE* 1218b37ff.

the non-fulfillment of which could explain, e.g. our being disappointed at the musician's performance. The point that the Strawsonian would make is that, although this negative response of ours involves a certain expectation or demand, it is not an expectation that could be analyzed in terms of emotionally toned attitudes like intrinsically moral disapprobation or 'deep' blame, or any other cousin of resentment.[70] What seems to be involved in the technical assessment is merely the belief that the runner's performance displays a good exercise of running capacities and the musician's performance a poor exercise of musical capacities, and that this reflects well or badly upon them as runner and musician respectively.[71]

Perhaps this point could be expressed differently, so as to let us see Aristotle's view on praise and blame in contrast with the more general notion of accountability. One could say that, on the accountability model, one is morally responsible for an action if and only if there is someone else who has a *right* to demand or expect from one a certain sort of behavior. My point could be captured by the claim that, just as we cannot be plausibly regarded as having a *right* to a good performance by the runner (unless the runner is bound to us by some contract, but this is a different matter), an Aristotelian ethical agent

---

[70] Writers in the Strawsonian tradition like Susan Wolf or J. R. Wallace talk about 'deep praise and blame', as opposed to 'superficial praise and blame', as a sort of praise that is inherently moral. See Wolf, S. (1990), pp. 40-3; Wallace, J. R. (1994), pp. 52-3.

[71] Sauvé-Meyer, however, has a retort to the foregoing argument: "One might worry" – she says – "that since Aristotle includes strength and speed among his examples of praiseworthy things, the praiseworthiness he here attributes to states of moral character is not specifically *moral*. But in *EE* viii 3, *Aristotle restricts the account of praiseworthiness to exclude such states,* on the grounds that the goods they produce are not good without qualification (*EE* 1248b19-24). He restricts praiseworthiness to the virtues of character because only their products cannot be misused. *Praiseworthiness in this restricted sense is a specifically moral assessment*." Sauvé-Meyer, S. (1993), p. 56, footnote 22 (my emphasis). I don't think her argument is convincing. The emphasis in *EE* VIII 3 is on the question '*How much* of the natural goods should one pursue, have, etc.? This question hardly makes sense as applied to *skills*. Aristotle in this context does not have in mind *skills*, but only *natural attributes* in so far as these are 'natural goods'. The *EE* VIII 3 claim that *natural goods* can be misused and that they are not praiseworthy in themselves, but only if employed in the right way by the virtuous person, does not have in scope the *ischuros* or the *dromikos* referred to at *EN* 1101b16, for these are not natural goods, (e.g. natural strength or speed). The victorious (*nikôntes*) who are *crowned*, e.g. in a contest like the Olympic games, at *EE* 1219b10, just like the *dromikos* who is e.g. *good at* running in the Olympic games, and the *ischuros* who is e.g. good at the Pankration or wrestling (*EN* 1101b12-18) are in possession of bodily *skills*, sharply distinguished by Aristotle from natural goods at *EN* III 5 1114a21-29, precisely on the grounds that they have been acquired through exercise and care. If Aristotle is not considering skills at all in *EN* VIII 3, then this chapter provides no evidence that, for Aristotle, skills are not praiseworthy in a way that is importantly similar to the way moral virtues are. Aristotle's position could be that we praise different things: in some contexts we praise people for musical skills, in other contexts we praise people for courage, etc. and the praise is the same (although the practical implications are not).

is such that other ethical agents do not have a right to demand a good performance from him – or rather, if they happen to have such a right, such right is not a necessary ingredient to an account of how and why, according to Aristotle, such a performance is attributable to him or to her.[72]

Some worries may arise from the fact there is an important disanalogy between an ethically bad performance and e.g. a musically bad performance, in Aristotle's own words: "in the matter of skills voluntary error is preferable to involuntary error, while in the matter of virtue the reverse is true" (*EN* 1140b22-23). Stendhal's voluntarily spelling a French phrase incorrectly in a lecture does not thereby make him a bad grammarian (suppose he doing it because he is putting to the test the grammar-level of his students), on the contrary, it is involuntary grammar mistakes that reflect badly upon him as a grammarian; in the case of ethical dispositions, on the contrary, it is voluntary mistakes that matter for the ethical evaluation of the ethical agent. That there is this difference between *the mistakes* that count as relevant for technical and ethical negative evaluation, however, does not impugn my claim that *the kind* of negative evaluation is in both cases the same: i.e. simply *psogos*. If the blame bestowed upon ethical agents were, on Aristotle's view, 'moral' blame in a substantive sense, this would be because it carries *in itself* the mark of morality, e.g. it would be S-reactively charged (or involve the belief that being S-reactively charged would be appropriate) or it would be the expression of a sanctioning disposition; but no such intrinsic mark of morality seems to be a necessary ingredient in the sort of blame peculiar to the domain of the technical, nor is it to be found in Aristotle's account of blame as directed at the ethical.

Now, if (i) this sort of technical assessment differs so radically from the sort of moral assessment identified by the Strawsonian persuasion and accountability in general; and if (ii) Aristotle is clearly not interested in singling out any sort of distinctively 'moral', 'deep' praise or blame, that is appropriate to virtuous or vicious agents and not to

---

[72] I am not assuming that the notion of accountability applies exclusively to actions that have breached *moral* demands. This is certainly not the case. Just think about cases in which people whose job (i.e. responsibility) is to perform a specific task is held accountable to some authority for doing it well or badly (e.g. fulfilling the conditions in the contract, etc). The point is that this is obviously not the sort of case relevant to Aristotle's examples of the victorious in a contest or the good runner.

technical agents, then this is of the utmost significance to our understanding of Aristotle's view on what is implied in our ascription of good or bad deeds to human agents: the fact that technical appraisals are, in this sense, closely parallel to ethical appraisals adequately captures Aristotle's perspective on the nature of the latter.

II

ARISTOTLE ON ETHICAL ASCRIPTION

THE POSITIVE ARGUMENT

In this chapter I develop a recognizable aspect of moral responsibility that I think correctly represents Aristotle's concerns with the conditions of voluntariness in his Ethics. I do this in various steps. First, I show that these concerns are best modelled on what Gary Watson calls 'attributability', as distinct from accountability. In the rest of the chapter I identify those features peculiar to Aristotle's own version of attributability, which I call 'ethical ascription' to distinguish it from 'Watsonian' attributability: (a) as distinct from Watsonian attribution of ethically significant actions to agents, Aristotelian ascription is meant to be overtly expressed; (b) it is also an attribution made from what I call 'the apt observer stance'; (c) it is based on the distinction between focal and derivative meaning of praise and blame; and (d) it applies to the domain of reason-responsive agents (as opposed to morally accountable agents on the one hand, and to pain-responsive agents on the other).

A. Accountability and attributability

The setting of a moral community in which persons (i.e. mutually accountable agents) require certain standards of conduct from each other, are disposed to respond unfavourably to another's voluntary failure to meet these shared demands, and their deserving these unfavourable responses for thus failing, is what characterizes the notion of accountability. According to accountability, B is responsible *to* (i.e. accountable or answerable *to*) A, and to everyone A represents, for complying or not with these demands or expectations.

What emerged from my arguments in Chapter 1 is that Aristotle's concern with the conditions of voluntariness in his Ethics is not primarily a concern with the conditions of moral responsibility as accountability. There was no reason to think that Aristotle's concept of the prohairetic agent is identical with our modern concept of the morally accountable agent (even if these two concepts have the same extension), nor was there any reason to think that the blameworthiness or praiseworthiness appropriate to human agents for their actions, can be analyzed in terms of the S-reactive attitudes or, more generally, in terms of sanctioning

dispositions (i.e. the praise and blame appropriate to them is not grounded on their breaching or meeting certain requirements of interpersonal relations, nor do they *deserve* or *merit* praise and blame for their ethically significant responses). Further, we found no convincing argument for thinking that children should be categorically excluded from the recipients of Aristotelian praise and blame, and lastly, we found positive reasons for thinking that the sort of praise and blame that Aristotle has in mind is sufficiently akin to the sort of assessment appropriate to skills and skillful activities, to warrant the suspicion that an altogether different model of what Aristotle means by 'S is blameworthy/praiseworthy for voluntarily doing A' is required. But what model?

Such a model is provided by Gary Watson's notion of attributability.[1] Watson argues that there is available another notion of 'responsibility' (his term) that can also be construed as *moral*, which does not correspond to, nor is parasitic on, the notion of accountability. [Of course, being the *cause* of one's actions is independent of people's demands or expectations with regard to one's actions (more emphatically, independent of owing obligations to one another), but this is an obvious and uninteresting point].[2] The interesting idea is that not all cases of being *morally* responsible for one's actions are analyzable in terms of *having* certain responsibilities or being *answerable* to certain people. There is a sense of 'moral' in which one can say, without invoking others' moral demands, that S is morally responsible for an action. Gary Watson offers a good illustration of this claim:

If someone betrays her ideals by choosing a dull but secure occupation in favour of a riskier but potentially more enriching one, or endangers something of deep importance to her life for trivial ends (by sleeping too little and drinking too much before important performances, for example), then she has acted badly – cowardly, self-indulgently, at least unwisely. But by these assessments we are not thereby *holding* her responsible, as distinct from holding her to be responsible. To do that, we would have to think that she is accountable to us and to others, whereas in many cases we suppose that such behaviour is 'nobody else's business'. Unless we think she is responsible to us or to others to live the best life she can – and that is a moral question – we do not think she is accountable here. If her

---

[1] Watson, G. (1996).
[2] It is uninteresting because this is the sense of 'responsibility' involved in the notion of agency, and in this sense an agent can causally responsible for an involuntary piece of behaviour.

timid or foolish behaviour also harms others, and thereby violates requirements of interpersonal relations, that is a different matter.[3]

Assessments of the sort here distinguished from accountability are made from 'the aretaic perspective'. Underlying this perspective, Watson argues, is one set of interests central to 'the ethical life', the one that "hinges upon our concern with living a good human life, with models and ideals of human possibility".[4] This 'face of moral responsibility', characterized by the aretaic perspective and the set of interests underlying it, is what Watson calls 'attributability'.

In contrast with accountability, which focuses on three elements – an agent A holding B accountable or answerable to certain moral requirements R – attributability focuses on two elements; the agent and his actions. What is important, from the point of view of attributability, is not the causal relationship between conduct and the agent (for such a relation might even be analyzed in non-causal terms), but the fact that such conduct is *expressive* of the agent's self, his values and adoption of ends, and that it is *executive* of his or her will. From this aretaic perspective, appraisals like "he acted in a cowardly way" can be understood, as expressing an appraisal of the agent *as coward*, that is, as implying that such cowardly conduct is *attributable* or *ascribable to a kind of fault in the agent*, i.e. an ethical fault or vice (which, by the way, does not have to be a fixed, stable character trait).

The blaming judgements involved in ascribing faulty actions to an agent's defect of character, when made from the aretaic perspective, are not appropriately described as 'sanctioning responses' or as expressing sanctioning dispositions. In a sense, of course, they are 'responses' or 'reactions' occasioned by someone's action, but they do not, as such, involve sanctioning dispositions.[5] Because of this, they are not appropriately characterized as responses that the agent 'deserves' (in virtue of his not conforming to the judger's normative expectations). Therefore, from the

---

[3] Watson, G. (1996), p. 231.

[4] *Ibid*, p. 240.

[5] The term 'disposition' here is very important. Someone could argue that the ethical condemnation of someone beyond reach of sanction, e.g. an historical figure, is thereby on the attributability side of Watson's distinction. This would be a hasty conclusion to draw, however. It still needs to be shown that such ethical condemnation does not express a *disposition* to sanction such person, e.g. perhaps one *would* sanction him if he were alive.

perspective of attributability, questions concerning the degree to which a certain piece of conduct deserves a certain response do not naturally arise. ('Desert' here is the narrow normative concept isolated in Chapter 1, Section F). Moreover, the attribution of an action to a fault or excellence in the agent does not have to imply that such fault or virtue is of an ethical nature, as opposed to a technical one, since there is nothing intrinsically 'moral' about aretaic appraisals, as there is about S-reactive attitudes.

Watson offers an interesting illustration of how the distinction between attributability and accountability can account for cases in which our intuitions as to whether an agent is responsible for a given action diverge. Think about the case (described by Watson) of the vicious criminal, who is not a psychopath, but who has been himself a victim of abusive childhood. "His deliberate and remorseless murders characterize him as malicious and cruel in a sense that no non-reflective being could be. The fact that life gave him a rotten deal, that his squalid circumstances made it overwhelmingly difficult to develop a respect for the standards to which we would hold him accountable, does not impugn these aretaic appraisals. His conduct is attributable to him as an exercise of his 'moral capacities' …"[6] On the other hand, Watson notes that "there is an inclination to doubt that such a person can rightly be *held accountable*, at least fully; that while he might 'deserve pity' … he 'does not deserve blame'." According to Watson, "what gives rise to our 'pity' are concerns about fairness. Facts about his formative years give rise to the thought that the individual has already suffered too much and that we too would probably have been morally ruined by such a childhood."[7]

In so far as to live the best life one can is a 'moral' question, then attributability is one aspect of 'moral responsibility', as Watson claims. But here 'moral' needs to be distinguished from the social network of demands or obligations, actions complying or failing to comply with them, and the corresponding

---

[6] Watson, G. (1996), p. 240.

[7] *Ibid*, p. 240. Philosophers who adopt a Strawsonian conception of moral responsibility of course agree with this. Peter Strawson himself includes seeing someone as "peculiarly unfortunate in his formative circumstances" (Strawson, P. F. (1962), p. 79) as the sort of agent-based consideration that tends to inhibit the S-reactive attitudes and to promote, instead, the objective attitude. See also Fischer & Ravizza (1998), p. 187; and Wolf, S. (1990), pp. 85-6.

institutionalized or informal sanctions or rewards (all of them elements of accountability). Accordingly I will keep the term 'ethical' to stand for the former interest and to keep 'moral' to stand for the later, and will call Aristotle's own version of attributability 'ethical ascription'. To the characteristics peculiar to ethical ascription I now turn.

B. Aristotle's own version of attributability

A consideration of Aristotle's view that praise is a *logos* suggests an interesting revision of modern, 'Watsonian' attributability – to coin one term for the occasion – in Aristotelian terms. Someone who "is afraid of the flies that flutter by him, cannot abstain from the lowest actions if they serve to gratify a desire to eat or drink, or loses his dearest friends for the sake of a farthing" (*Pol*. 1323a32), would be judged as 'cowardly' or 'intemperate'. According to Watson, to appraise a person in these terms is not *eo ipso* to hold him accountable to us and to others, and we should not regard this appraisal as necessarily involving a disposition to sanctioning behaviour. The difference between Aristotle and a modern 'attributivist' (such as Watson) is that the latter would also describe the aforementioned blameworthy conduct as one that is 'nobody else's business' (given that the person on the other end is a sane adult), whereas Aristotle with all probability would not describe it as such. A member of a modern liberal democracy may think he is morally obliged to keep his negative appraisal to himself (some remarks of Mill's *On Liberty* illustrate this), but not a citizen of an Aristotelian *polis*. As we have seen, Aristotle seems to think that praising and blaming are prominently judgements *in words* or *dictions* (*lexeis*, *Rhet*. 1367b36-1368a9) of people's ethical quality or their actions'.

What underpins this view of praising and blaming *logoi*, I surmise, is the function of the Aristotelian *politikos* and the nature of the *polis*. The main *telos* of an Aristotelian city-state is the good life (*Pol*. 1278b22-24), of which excellence and its exercise is an essential ingredient (1283a25). The best *politeia* is the one under which "anybody whatsoever would be best off and would live in happiness" (1324a23-25), where living in happiness is living the life of excellence (1328a38). This, of course, led Aristotle to the anti-liberal belief that the *politikos* must be

chiefly concerned with the cultivation, involving both education and upkeep, of virtuous citizens (*EN* 1099b29-23); hence the description of the Ethics as an investigation of a political nature (*politikê tis*, *EN* 1094b11). Accordingly, it would not be unreasonable for Aristotle to hold that the citizens of a *polis* are bound by the duty to overtly express a concern about each other's excellences and defects of character, regardless of whether these are directly harmful or beneficial to others. This is perhaps the reason why Aristotle does not seem to think that praise is an approving judgement, which can be held privately. It is reasonable to think that, consistently with the Athenian tradition, it was a civic duty for Aristotle to celebrate excellence.[8]

Despite the overwhelming emphasis the Greeks placed on praise, the publicity of blame should also be acknowledged. It is reasonable to think that Aristotle's silence with regard to the nature of blame is only due to the special weight he gives to theorizing about excellence, and not to the fact he is only interested on the ascription of good actions, for he is obviously not. Besides, there is no good reason to think that the *politikos*' concern with promoting virtue amongst citizens would not also benefit from the civic duty to openly impute bad deeds to the bad person. Indeed, it is because praise and blame are equipollent in this regard that Aristotle omits a discussion of the nature of blame in the *Ars Rhetorica*: how to blame (where this of course includes, amongst other things, blaming qualities of character) is something we can infer from what has been said about how to praise (1368a8-9, a36-37). It is then reasonable to think that the same conclusion that we have reached in the last paragraph with regard to praise is also pertinent to blame.

But herein lies the problem. Even if public ascription of good deeds to good people can be understood in terms of Watson's notion of attributability, it remains difficult to see – at least for us liberal moderns – how is it possible for a public and overt ascription of *bad* deeds to a person not to be understood as an expression of

---

[8] As Dover notes, public and institutionalized reward for virtue, like the award of a crown, and commendation formally expressed by the assembly and even recorded in an inscription, was a common practice in the 5th Century BD Athens. Dover, K. J. (1994), pp. 229-30. Indeed, the reason why Aristotle is interested in the nature of praise in the *Ars Rhetorica* is that he is interested in a particular branch of rhetoric, the epideictic (*epideiktikon*) or 'demonstrative' (as distinguished from the judicial and deliberative branches), which is particularly concerned with praise (and blame) as manifested in *public* speeches (see *Rhet.* I 9).

sanctioning dispositions (ranging from resentment or moral indignation, to punishment, or the belief that any of these would be appropriate). One can show, however, that at least at a theoretical level there is no conflict in thinking of an overt, non-instrumentalizing attribution of blameworthy behaviour, as something else than an expression of sanctioning dispositions, and thus to conclude from this that Aristotle's ethical ascription remains on the attributability side of Watson's accountability/attributability distinction.

One interesting way of doing this is by highlighting the role of *shame* (*aidôs*) in Greek culture, a culture that Bernard Williams, borrowing the expression from E. R. Dodds, called a 'shame culture'.[9] Although Strawson includes shame (a self-regarding reactive attitude) amongst the S-reactive attitudes, it has been suggested that this was a mistake of his. From the point of view of the recipient of blame, the standard contrast appealed to is the one between guilt (a genuine self-regarding reactive attitude) and shame. Whereas guilt is said by some philosophers to involve "an acknowledgement of one's blameworthiness that recognizes both the grounds of blame and ... the authority to level it", shame involves seeing oneself as "an object of the other's regard or 'gaze'", thus corresponding to the adoption of a "purely 'objective' view of oneself".[10] It is noted, for instance, that whereas guilt prompts one to respond in the appropriate ways to the corresponding impersonal reactive attitude (e.g. by apologizing, making the necessary amendments, or perhaps simply by accepting the sanction), shame "inhibits second-personal engagement – one feels like escaping from view".[11] The question then is: what sort of impersonal or other-regarding 'attitude' does the sense of shame correspond to? Whatever the other-regarding 'attitudes' corresponding to the sense of shame are – Williams suggests contempt, derision and avoidance[12] – Aristotle makes it clear that they involve at least an *opinion* (*doxa*) about the person prone to feel ashamed (*EE* 1233b27, *EN* 1128b24, *Rhet*. 1384a14), and opinion is all we need to make sense of a sort of overt blame that is not the expression of a sanctioning disposition. The key point is that, since shame does not require a corresponding interpersonal S-reactive attitude like

---

[9] See Williams, B. (1993), Chapter IV.
[10] Darwall, S. (2006), p. 71.
[11] *Ibid*, p. 72. See also Williams, B. (1993), pp. 88-95 and his Endnote I.
[12] Williams, B. (1993), p. 90.

moral indignation, it seems that 'the shame-system' can provide us with a framework within which a non-punitive, non-reactive mode of public blame is possible.

To see how this suggestion is embodied in Aristotle's Ethics, consider the operation of the Aristotelian law. The sort of blame that appeals to the sense of shame is most likely to be embodied in the law, as Plato and Aristotle view it. If the *politikos* is to promote excellence amongst citizens, then the law is likely to be extremely useful in helping him to succeed in this enterprise, and if the law is to help the statesman in this enterprise, it had better not address *well disposed* citizens (i.e. citizens prone to feel ashamed) merely by announcing punishments for lawbreakers, for if laws are felt to appeal merely to the well-disposed young citizen's fear of the sanctions incurred as a result of breaching the law, then the law will not be respected for its own sake (*EN* 1179b11-13), thereby losing its power to educate well-disposed young citizens in the ways of virtue. This is something Aristotle himself recognizes in *EN* X 9. The law will address A – if A is a well-disposed young citizen – in ways that will appeal to A's rational nature, by "encouraging him to virtue and persuading him in the name of the fine" (1180a6-7). How? Well, by appealing to A's sense of shame. Aristotle assumes that the law will address young citizens by means of persuasion and exhortation only if they "by nature obey their sense of shame" (1179b11). He thinks (*EN* IV 9) that shame is characteristic of young people, and it is their being prone to shame what makes them the appropriate targets of the 'persuasive' and 'educative' face of the law. The purely punitive, sanctioning aspect of the law's abiding force becomes then secondary for the purposes of addressing well-disposed citizens, and so the Aristotelian law furnishes a concrete example of a sort of overt, indeed institutionalized, blame that is not necessarily the expression of sanctioning dispositions.[13] Incidentally, it is not insignificant for my overall argument that, as Bernard Williams notes, shame – unlike guilt – is neutral on the moral/non-

---

[13] It has been suggested to me by Sarah Broadie that Aristophanes' plays can also provide us with examples of non-punitive, overt blame. These plays are full of characters (which in many cases are real people) that are mocked on the occasion of their 'inoffensive' vices, providing us with a large array of examples in which openly imputing vicious deeds to someone does not necessarily involve dispositions to sanctioning behavior.

moral distinction: "we, like the Greeks, can be as mortified or disgraced by a failure in prowess or cunning as by a failure of generosity or loyalty".[14]

C. The question about Aristotle's concern with voluntariness: focal and derivative praise and blame

Let us now return to the main question raised at the beginning of Chapter 1, Section A, about Aristotle's concern with the conditions of voluntariness in his Ethics. There we saw that in the *EE* and *EN* passages where Aristotle justifies his concern with these conditions (*EE* 1223a9-15; *EN* 1109b30-34), he makes the following two claims:

(i) Virtue and vice have to do with (*peri*, *EN* 1109b30; *EE* 1223a14) voluntary responses; and

(ii) Voluntary responses, and not involuntary ones, are praised and blamed (*EN* 1109b31; *EE* 1223a9-13).

Now, we also saw that the relation between virtue and vice and voluntary responses, captured by '*peri*' in (i), was flexible enough to encompass both (A) the causal relationship between particular ethical dispositions or adult agents in possession of them and the virtuous or faulty responses they are disposed to have; and (B) the genetic relationship between certain voluntary responses the frequent practice of which generates excellent or faulty ethical dispositions. We also saw that this ambiguity of 'virtuous activity' captured by (A) and (B) is promoted by Aristotle's belief that both the *actions* involved in character formation and the ones expressive of an already possessed character are of the same in kind (*EN* 1114b26-27).

In a series of passages where Aristotle highlights the significance of voluntary actions for praise and blame, however, he appears to focus exclusively on (A), that is, on those ethically significant, voluntary responses that come about as a result of the

---

[14] Williams, B. (1993), p. 92.

exercise of an excellent disposition. Aristotle says in the *EN* that "we praise the just or brave man and in general (*holôs*) both the good man and excellence itself *through the actions and deeds involved* (*dia tas praxeis kai ta erga*)" (*EN* 1101b14-16) and later that "praise is appropriate to excellence (*epainos tês aretês*), since *as a result* of excellence people are disposed to do fine actions (*praktikoi tôn kalôn apo tautês*)" (*EN* 1101b31-32). Again, in the *EE* he says that "excellences and defects and the actions *that spring from them* (*ap' autôn erga*) are praiseworthy and blameworthy respectively" (1223a9-10), and the *MM* corroborates these claims by stating that "excellences are praiseworthy things (*epanieta*): for praise originates from the actions that are *done in accordance with them* (*apo tôn kat' autas praxeôn ho epainos ginetai*)" (*MM* 1183b26-27).

Having rejected the Strawsonian rationale for an exclusive concern with sane rational adults and their conduct, we can now offer an account of praise and blame that makes room for the whole range of responses virtue and vice are 'concerned with' (*peri*, *EN* 1109b30; *EE* 1223a14), as both the voluntary responses involved in character-acquisition and those that are the full manifestation of character traits. At the same time, this account must accommodate these passages in which Aristotle seems to be exclusively concerned with conduct fully expressive of ethical disposition. This is the aim of the present section.

In order to do this, it is crucial to start by drawing a distinction between *formal* and *material* praise and blame. (1) It is one thing to claim that 'praise is of courage' and 'blame is of cowardice' or that 'praise is of the courageous' and 'blame is of the cowardly'. These claims are *meta-claims* about praise: as we have seen in Chapter 1 (Section E), they mean that terms like 'courage', 'cowardice', etc. have a laudatory and pejorative meaning respectively (or – alternatively – that *qualities* like courage or cowardice have value or disvalue). (2) Something very different, however, is calling an *individual* like Alexander or Coriscus 'courageous' or 'intemperate' ('praise is of Alexander'), which is already a way of praising or blaming *him*: this is praise or blame. Here the point is material, in the traditional sense that it provides us with new information about the subject.

This distinction is important, because in the passages just quoted Aristotle has in mind meta-claims about praise and blame, as described in (1) above: this is

suggested by his talking about excellences and faults of character as *qualities*, as opposed to excellent or faulty *individuals* like Alexander or Coriscus. But how, then, does Aristotle's point about *actions* (i.e. praise is 'of' excellence 'through' (*dia*) or 'from' (*apo*) deeds) in these passages contribute to making these meta-points about praise and blame? I think that the claims just quoted supra (seen as meta-claims) maintain that *we value* virtue *as a quality* (or the ideal of the virtuous agent, the *agathos*) because it *leads* to, or is the source of, good and noble achievements. These claims *imply* that we value an excellent *individual* to the extent that he *has* virtue – and, as we know, virtue makes him *praktikos* of fine deeds – but it is not what they mean.[15]

(These claims about the value of virtue as a quality are perfectly compatible with the claim that 'praise is of virtue' should be understood as 'virtue-terms are laudatory', rather than as 'praise is of something already independently identifiable as a virtue' - indeed, they could be taken as providing an account of *why* virtue-terms are laudatory, e.g. they are laudatory, because the virtuous qualities themselves (the extra linguistic entities virtue-terms refer to) are ultimately valuable as being the source of fine achievements. The same goes, *mutatis mutandis*, for the disvalue of vice as a quality). Now, of course, virtue and vice *as qualities* can be the 'sources' of noble deeds, either because the agent in possession of them is such as to do noble deeds, or because they function as the ultimate goals of moral education.

Now, *because* virtue- and vice-terms are laudatory and pejorative in meaning, or because virtue- or vice-qualities have value or disvalue respectively

---

[15] Also important is to note that these claims do not have to imply that the achievements or deeds in question (*erga*) are independently identifiable as 'good' or 'fine', in such a way that their goodness or nobility 'reflects back' to their causal source, i.e. the ethical excellences. Rather, good and noble *erga* are the virtue itself *in activity*. Thus, Aristotle does not have to imply that the value of virtue is derivative from the value of those good and noble deeds of which virtue is 'productive' (as the value of craftsmanship is derivative from the independent value of its products). Here, it is still true that logical primacy belongs to the actual behaviour, but this is because of the primacy of *hexis* in activity or in actuality (*en energeiai*) over *hexis* merely in potentiality (*en dunamei*), which is not the relation of distinct product to cause. All Aristotle needs to imply is that good activity is more valuable than good potentiality, and indeed he argues for this metaphysical thesis *in extenso* (*Met*. IX 9; *EE* 1219a14-18). Since such good and noble deeds *are* the virtue in activity, then valuing virtue in general because it leads to good and noble achievements amounts to valuing virtue itself *in activity*. On my interpretation, the sort of relation that holds between an ethical disposition and its *erga* is the *dunamis-energeia* relation as 'ways of being': *one and the same thing* can either *be* or be considered in two ways, *kath' hexin* or *kat' energeian*, in relation to its mere dispositional existence or to its actual existence (e.g. *EN* 1098a6, 1157b5-11, 1167b30-1168a9).

(meta-claims), *calling* (a) an individual like Alexander or Coriscus 'courageous' or 'self-indulgent' on the occasion of his recognizably courageous or self-indulgent conduct, or *calling* (b) that conduct itself 'courageous' or 'self-indulgent', are *ways of* praising or blaming *him* (material point). The crucial point here is that (a) calling an individual 'courageous' or 'self-indulgent' does not have to imply that he is already in possession of a just character. When in the process of moral education I praise a child by calling him 'just', this assertion may take the form of saying that he is *on the way to* being a just person, and perhaps also of encouraging him to keep behaving in the same way as he is now acting (i.e. to keep doing just actions). Aristotle, for example, denies that animal conduct can be correctly designated by virtue- or vice-terms (*EN* 114525-7; 1149b31-5), but he never denies this of children or people in the process of moral reform. And he is certainly right in not doing so, for, as I will argue in Section E, whereas children are reason-responsive, animals are not.[16] Likewise, it is not at all clear that (b) calling a piece of behaviour 'courageous' or 'self-indulgent', necessarily invokes *full manifestation* of the corresponding virtue or vice instead of, for instance, acting *like* a courageous or self-indulgent person. As a matter of fact, Aristotle makes this latter point explicit in *EN* II 4, when he invokes the notion of the ideal just or moderate agent as a criterion for determining which *actions* can be called 'just' or 'moderate':

C: Things done are called 'just' and 'moderate' (*ta pragmata dikaia kai sôphrona legetai*) whenever they are *such that the just person or the moderate person would do them*. *EN* 1105b5-7

Accordingly, Aristotle's insight is that while rational *individuals* (either whether mature or immature) as well as ethically significant *actions* can be subjects of ethical appraisal in virtue- and vice-terms, these terms refer primarily to ethical dispositions *as qualities* (as specified by our notion of formal praise and blame). The claims I have made thus far can be nicely captured by recourse to the notion of *focal*

---

[16] In fact, when Aristotle recognizes that "sometimes we praise people who are deficient (*tous elleipontas*) and say that they are good-tempered, and sometimes those who are angry, calling them manly" (1109b16-18), he is recognizing that people who fall short to a certain extent of hitting the right mean in their actions are still praiseworthy, and there seems to be no reason to exclude would-be moral agents or agents in the process of moral reform from this group.

meaning, coined by G. E. L. Owen to render the Greek '*pros hen*'.[17] The concept of focal meaning applies whenever there is a semantic connection between the various uses of a word in virtue of one primary use, in terms of which the rest of the uses can be explained. To use an example offered by Aristotle himself in *Met*. IV 2, the focal meaning of 'healthy' is the one that refers to a living body in a good condition. Thus, we can say that an apple is healthy, or that a given complexion, or that an activity is healthy, but these further applications of 'healthy' are explained in terms of their relation to the focal one: the apple is healthy because it helps keeping the body in a healthy condition, and a given complexion is healthy because it is an indication that the body is in a good condition.

We can now apply this distinction between focal and derivative meaning to virtue and vice-terms, and say that in their focal use virtue- and vice-adjectives refer to ethical dispositions as qualities (or to the ideal of the fully developed agent in possession of an ethical disposition). In their derivative use, virtue- and vice-terms are applied (I) *to actions* in virtue of a certain relation that actions have with this focal use. This relation is twofold, as the *Ars Rhetorica* makes clear: Virtue is the *epaineton* ('the praiseworthy') par excellence (1366a33-36); accordingly, whatever stands to virtue in the relation of tending towards virtue (*pros aretên*, 1366b25) must be *epaineton* - these are the things that produce virtue (*ta poiêtika tês aretês*) - and whatever stands to virtue in the relation of being its signs or works (*sêmeia ê erga*, 1366b29) also must be *epaineton* - these are the things that come from virtue (*ta ap' aretês ginomena*).[18] Both types of actions have in common their being such as the ideal agent in possession of an ethical disposition would do them (i.e. C *supra*): these are voluntary actions that can be characterized as 'what the V-agent would do' (as Aristotle makes it clear, when introducing criterion C *supra*).[19] Moreover, in their derivative use, these evaluative adjectives can be applied (II) *to agents*: both to agents in full possession of ethical dispositions, as well as to ethically immature agents like children or agents in the process of moral reform. In this use of

---

[17] Owen, G. E. L. (1960).
[18] See also *De Virt.* 1249a26-30.
[19] When Aristotle states criterion C *supra* (*EN* II 4) he makes it clear that by 'just' (*dikaia*) or 'moderate' (*sôphrona*) *pragmata* (1105b5-6) he means voluntary actions, for these *pragmata* have just been characterized as the actions the repetition of which generate a stable character (b4-5), and these are voluntary actions (*EN* 1113b26-27; 1114a12-16).

'just' or 'moderate', calling a child 'just' is making a general reference to just character or to the ideal of the mature agent in possession of a just character, but *not* to the child's current character.[20]

Now, if calling a person or his actions 'just', 'courageous', 'moderate', etc., happens to pick out a fully mature, prohairetic agent and that conduct which is fully expressive of his developed ethical disposition, then these terms describe *him* and his conduct as they actually *are*, but the fact that praising him in these terms succeeds in so describing him and his conduct is not in itself attributable to the nature of praising and blaming *logoi*, in the way in which the success of S-reactive praising or blaming attitudes in picking out a morally accountable agent is – on the accountability model – attributable to the very nature of these attitudes, i.e. it is what these attitudes 'aim at', so to say. It is much more plausible to think that the message conveyed by praising and blaming *logoi* in both their focal and derivative applications is *the same*: 'keep/stop acting/being in the way the V-person acts/is'.

Praising an action as 'just' then is like saying 'that's it, that is exactly how the just person conducts himself in this particular situation', and the same goes for praising a child as 'just'. In fact, praising or blaming an action as 'v' is just a way of praising or blaming its agent as 'V'. *Individuals* are praised and blamed *through* (*dia*) the praising and blaming of their actions, and responses in general. I take it that in the few places where Aristotle is making a *material* point by emphasizing the role of action in praise and blame (e.g. *EE* 1223a11-12), the primary force of the '*dia*' is precisely to mark those actions *through which* individuals are praised: i.e. calling an action 'v' (according to criterion C) is a way of calling its author 'V'.

Let me then say that Aristotle is concerned with the conditions of voluntariness in his Ethics, because he is concerned with the following question:

Qe: What are the conditions that need to be fulfilled in order for an apt observer to be warranted in (1) praising and blaming an action A as 'v' (on the basis of criterion C

---

[20] I take it that this is another way to express R. M. Hare's idea that "sometimes we commend an act within the class of acts indicative of his [the agent's] moral character, that is to say, those acts which show whether or not he is a good man – and the class of comparison 'man' in this context is the class 'man to try to become like.'" Hare, R. M. (1952), p. 144. Hare himself refers to Aristotle's remark that 'a good act is the sort of act that a good man would do', p. 186. On this see also Broadie, S. (1990), pp. 127-8.

supra) and (2) praising or blaming a reason-responsive agent S in 'V'-terms through (*dia*) (1)?

And let me call Aristotle's concern with these conditions, a concern with 'the conditions of praise and blame' or, alternatively, with 'the conditions of ethical ascription'. Notice that I have introduced in this question the concepts of the apt observer and the reason-responsive agent; I explain these two notions in the following two sections. An action is blameworthy or praiseworthy (on the basis of C) only if voluntary. The conditions of praise and blame then include the conditions of voluntariness. But there is more to this, as explained in the *Intro.*; for Aristotle also thinks that praising and blaming an action A as 'v' is a way of praising or blaming its agent S only if S had done A without the relevant sort of pain. This is a separate condition (i.e. (2) in Qe), because A's being done without the relevant sort of pain is best understood as a condition that is not entailed by A's voluntariness, as I argue in Chapter 7.

Finally, given that A's voluntariness is usually *assumed*, this allows for the possibility that the apt observer in question is in fact mistaken in making this assumption. Thus the mention of 'warrant' in Qe. I will also take for granted, as explained in the *Intro.*, that the assumption of voluntariness corresponds to the negative assumption that none of the factors defeating ethical ascription, i.e. neither the ones affecting the voluntary status of the action (e.g. force, factual error) nor the one affecting the voluntary status of the agent (e.g. moral pain), has been present. If this is the case, then both conditions, (1) and (2), are met. Finally, the concept of an observer being 'warranted' in making these two logically distinguishable moves is to be contrasted with the idea that the agent or his responses 'deserve' or 'merit' praise and blame, as accountability would have it.

D. The apt observer stance

I take it that an immediately appealing feature of Qe is that it treats the observer stance as central. This is important, for now it is a requirement that the conditions of praise and blame be verifiable by an observer. This seems to be the

case with the conditions Aristotle has in mind (e.g. absence of violence and knowledge of the particular circumstances of the action). Indeed, it seems to be a very appealing feature of these conditions that they are at least in principle *accessible* to an observer in a variety of ways, e.g. questioning (although they might not be *directly* observable, except perhaps the violence condition). This is an appealing feature precisely because, according to the long-lasting medieval dogma concerning voluntary action, a voluntary overt action is defined as being the effect of a mental act of willing or 'volition' to perform the overt action. This theory was justly criticized by Gilbert Ryle, partly because mental acts of willing are unverifiable even to the very mind engaged in them.[21]

Now, admittedly, the task carried out by an observer, of praising or blaming an agent S as 'V' through (*dia*) praising or blaming his action A as 'v' on the basis of criterion C, may seem virtually impossible. The reason for this is that a factual description of A alone will not be enough to establish A as *that which the V-person would do*; so that ethical ascription is in danger of being obstructed from the very outset. A factual description of A is not enough to establish this, because (i) such characterization also requires attending to the notion of opportunity or right moment (*ho kairos*, 1104a6), that is to say, the particular circumstances in which A occurs, or, as Aristotle puts it, the particular way (*pôs*) in which A is performed (*EN* 1106a1); the "what, when, why, how, whom, etc." in the Aristotelian formula of the mean (*EN* 1104b22-6; 1106b21-3); and (ii) such ethical opportunity or moment involved in the practical considerations of the ethical agent, because of its particularity, is not susceptible of falling under any set of non-empty rules or precepts (*EN* 1104a7). All this, of course, is also applicable to the precepts by which we *judge* other people's actions. Now, if recognizing A as 'just' or 'cowardly' is tantamount to recognizing A as 'what the just or cowardly agent would do', and if recognizing A under this description requires one to attend to the particular circumstances in which A occurs – so that the description above is just an ellipsis for 'the action that the V-agent would do *in these circumstances*' – then it does not only look as if the observer needs to fulfill a very demanding task just in order to start

---

[21] See Ryle, G. (1949), Chapter 2.

with ethical ascription; it also looks as if the observer needs to be *himself* a just or brave person, for otherwise how could he be capable of recognizing the action as 'just' in the full specificity of the particular circumstances?

This latter, I think, is not an unwelcome conclusion by Aristotle's own, very demanding standards imposed on the assessments involved in ethical ascription. The reason for this is that, according to Aristotle, the proper, impartial judge for all ethical behaviour is the virtuous person. An important feature of Aristotle's doctrine of the mean is that two *equally rational* ethical agents can see the same practical situation as demanding diametrically opposite sorts of action, in so far as they are located in different positions within the spectrum of ethical character (e.g. coward – courageous – rash) and thus judge of each other on the basis of that location. So for instance Aristotle says that:

Just as what is equal is larger when compared with the smaller and smaller when compared with the larger, so the intermediate dispositions are excessive when compared with the deficient ones and deficient when compared with the excessive ones [e.g. rashness], in the sphere both of affections and actions. For the courageous appear rash in comparison with cowards, but cowardly in comparison with the rash; similarly the moderate, too, appear self-indulgent in comparison with the 'insensate', but 'insensate' in comparison with the self-indulgent, and the open-handed appear wasteful in comparison with the avaricious, but avaricious in comparison with the wasteful. This is why those at either extreme try to distance themselves from the one between them, associating him with the other extreme: the courageous person is called 'rash' by the coward, 'cowardly' by the rash person, and analogously in the other cases. *EN* 1108b15-27[22]

Aristotle's point here is that faulty characters at the opposite extremes of the same spectrum display a radical disagreement as to how to characterize the virtuous lying at the centre, for they are prone to associate the virtuous person with their diametrical opposite: the cowardly sees the courageous as rash, and the rash sees him as cowardly. Because of this, faulty agents lying at the extremes will also attach the wrong evaluative description to their diametrical opposite (e.g. the coward thinks that courage *is* rashness). (How these agents characterize themselves, Aristotle does not say).

---

[22] Rowe's translation, with slight modifications.

Because their position in the ethical spectrum gives them a distorted perspective, not only of virtue, but also of the opposite vice, the evaluative judgement of faulty agents is deeply distorted. But having a distorted perspective in this sense does not give a sufficient condition for obstruction of objective ethical assessment; the sufficient condition seems to be provided by the fact that the person at the extreme in an ethical spectrum not only sees things askew, but is also unaware that he does, and cannot, accordingly, correct for it. His evaluative perspective is radically distorted, because he cannot switch between his perspective and the other's, and so his evaluative judgment is not only distorted: he is also unable to subtract what contributes to the distortion of his judgement. On the contrary, objectivity of ethical assessment, it may seem, is confined to the judgements of the virtuous person. As Aristotle says, "the good man judges each class of things rightly, and the truth is what appears true to him" (*EN* 1113a30; see also 1173b23-26 and 1176a17). This is then an important aspect of the familiar Aristotelian doctrine that the virtuous person is the *metron* (the measure) of each thing: only now as the person whose praise and blame is objectively right. Hence, I will be speaking of the *virtuous* or *apt spectator's* praise or blame.

E. Reason-responsiveness

Let me bring to an end my account of ethical ascription, by saying something about the *sort* of agents who are, on Aristotle's view, the appropriate recipients of praise and blame. These are the recipients who are able to absorb the focal message of praise and blame *logoi*. This message is basically, in the case of praise, 'act/be as the good agent acts/is', which of course applies to the good agent himself, in the form of '*keep* being a good agent and *keep* acting well'; and in the case of blame, 'don't act/be like bad person acts/is'. Recall the comparison between praising *logoi* and suggestions in *Rhet*. 1367b36-1368a9: if you want to praise, look at what you would suggest or advise.

My proposal in this section is that the sort of agents who can absorb this sort of message, are agents having an *ethical condition* (*êthos*). The Eudemian definition of *êthos* is particularly suitable for embracing the broad range of agency I have in

mind: *êthos* is "a quality (*poiotês*) of the soul that is *capable* (*dunamenê*)[23] of following reason in accordance with governing reason (*kata epitaktikon logon*)" (1220b5-7). Aristotle here defines '*êthos*' by recourse to the notion of reason-responsiveness, that is, agents in an ethical condition are *capable* of responding to reason. Reason-responsiveness provides us with the right criterion to distinguish those agents who are the appropriate targets of Aristotelian praise and blame, from those who are not. As we suggested in Chapter 1, both the Strawonian interpretation and the prospective conditioning interpretation set the boundary between the 'appropriate' targets and 'non-appropriate' targets in the wrong place. The first one sets this 'moral' boundary between Strawsonian, morally accountable or answerable agents, and non-accountable agents like animals, psychopaths and children. On my interpretation, this view is too restrictive because it leaves children out, and children are reason-responsive. The prospective conditioning interpretation, on the other hand, sets the relevant boundary between those beings that are responsive *to pain* and can be conditioned by chastisement, and those that are not pain-responsive. On my interpretation, this view is too inclusive, because it includes animals and psychopaths, and neither animals nor psychopaths are reason-responsive. My interpretation sets the boundary between the appropriate and inappropriate targets of praise and blame, in the distinction between reason-responsive agents (starting with children) and those agents that are not reason-responsive (like animals). This interpretation is strongly corroborated by the reason explicitly offered by Aristotle to withhold virtue- and vice-terms from *animals* and *insane* people: none of these agents have (*echein*) rational choice or reasoning (*EN* 1149b34-35).[24] As we saw in Chapter 1 (Section C), children do have (*echousin*) reason, though this is not fully present in them.

---

[23] I follow here Spengel's conjecture and Ross' excision of '*de*' at b6.

[24] Sauvé-Meyer claims that Aristotle "ascribes voluntariness to animals and children (*EN* 1111b8-9), but denies that their behaviour is properly classified in moral terms, as for example, temperate, brave, cowardly or just (*EN* 1145a25-7, 1149b31-5)". Sauvé-Meyer, S. (1993), p. 52. This is certainly not true of children. Not only is any mention of children absent from *EN* 1145a25-7 and 1149b31-5 – and my account explains why – but, on the contrary, Aristotle is committed to claiming that their behaviour is properly classified in moral terms, because (1) he thinks that criterion C *supra* (*EN* II 4) applies to voluntary actions, (2) he makes it clear that by 'just' (*dikaia*) or 'moderate' (*sôphrona*) *pragmata* (1105b5-6) he has in mind the voluntary actions the repetition of which generate a stable character (b4-5), (3) and these are obviously actions performed by children or people in the process of moral reform.

But what exactly is reason-responsiveness? Aristotle's most elaborate account of this notion can be found in *EN* I 13 (1102b13-1103a3).[25] In this chapter Aristotle introduces the taxonomy of the *psuchê* that will underpin his account of excellence of character (*ethikê aretê*). According to Aristotle (*EE* 1220a10-11) excellence of character is the excellence of the desiring part of the soul (*orektikon*), the seat of the emotions and appetitive desires. As Plato did in the *Republic*, Aristotle argues that this part or aspect of the soul is different from the intrinsically rational part of the soul, by pointing to the phenomena of continence and incontinence (1102b13-25): continent and incontinent agents exhibit, in their own peculiar ways, a *conflict* between what reason prescribes and what they desire. This something else (*allo ti*) that not only in these agents, but also by nature, opposes and fights reason, is the non-rational, desiring part of the soul. Now, essential to Aristotle's ethical project is to argue that the desiring part also shares in reason (*logou metechein*); in other words, to argue that it is reason-responsive, in spite of its recalcitrant nature. This is essential because what is peculiar to the *ethical* excellences is that they can come about through right habituation (*ethismos*), which involves the education of the desiring part, and these processes require reason-responsiveness.

Now, interestingly enough, reason-responsiveness has many stages of development, and each of these stages accounts for each of the main ethical conditions that Aristotle recognizes. Aristotle is very keen on illuminating this elusive *internal* relationship (i.e. internal to the agent) between the desiring part of the soul and practical reason by recourse to inter-personal attitudes or dispositions typically held towards an authority. Thus, he says that the desiring part is capable of *listening* to reason (*katêkoon*, *akoustikon*), that it is capable of *obeying* reason (*peitharchikon*), or that it is capable of *being persuaded*, in a way, by reason (*peithetai pôs hupo logou*). Reason, on the other hand, is represented as what prescribes (*epitaktikon*, *EE* 1219b30, 1220a9). Now, the elucidation of the internal relationship between desire and reason within an agent by recourse to these inter-personal attitudes or dispositions is not, as Aristotle sees it, a mere metaphorical

---

[25] See also *EE* 1219b26-1220a4; *EN* 1149a25-b2; *Pol.* I 5.

device. The relation between the inter-personal dispositions and their internal counterparts is much stronger than mere resemblance: it is genetic.[26] The interpersonal attitudes and dispositions involved in the willingness to listen, obedience, and openness to rational persuasion, are indeed the attitudes expected from a pupil in the process of moral education and right habituation. If ethical excellences, and ethical dispositions in general, are indeed the final products of these processes, then it is only to be expected that the *internal* relationship between desire and reason as exhibited in these developed dispositions, is the internalization of their inter-personal ancestors (*EN* 1119b13-15, 1138b7-9).

The most interesting advantage of our appeal to reason-responsiveness is that this concept shows that, in Aristotle's view, there is a continuum between the child and the fully virtuous agent: both are reason-responsive agents, although in different degrees. The proper *qualitative* difference between the Aristotelian 'appropriate target' of praise and blame *logoi* and the 'inappropriate target' is rather the different between the non-reason responsive agent like animals and people in beastly conditions (*EN* VII 5), and the reason-responsive one.

Thus, at the most basic level, reason-responsiveness is the disposition of the pupil (or the desiring part of the pupil's soul) to listen to and obey the orders of the parental figure. Accordingly, Aristotle declares that the desiring part can be said to *have reason* (*echein logon*)

In the sense of *echein logon* in which we say that someone who listens to or obeys his parents or loved ones 'has reason', and not in the sense of 'having an account', as in mathematics. That the non-rational is in a way persuaded by reason (*peithetai pôs hupo logou*), is indicated by the practice of admonishing people (*hê nouthetêsis*), and all the different forms of reproof (*epitimêsis*) and exhortation (*paraklêsis*). And if this [part] too is to be said to 'have reason' (*logon echein*), what has reason (*to logon echon*) too will be two-fold: in one sense, it is [what has reason] primarily and in itself, and in another as something capable of listening (*akoustikon ti*) as if to one's father. (1102b31-1103a3).

---

[26] As Freud observes, this is precisely the interpersonal relation between father and child and latter in the mature individual the super-ego/ego internal relationship. See Freud, S. (1991), Lecture 31 ('The Dissection of the Psychical Personality').

Notice that reason-responsiveness, as it is already exhibited by sane children, is the natural *capacity* of the desiring part of the soul to follow reason, not just the exercise of this capacity.[27]

Now, besides this primitive stage of reason-responsiveness *qua* natural, still untrained, capacity to obey or listen to external reason, there is another stage of reason-responsiveness, as found in the incontinent. Even though the incontinent, as such, acts contrary to reason, Aristotle gives every indication (although not conspicuously in *EN* I 13) that the incontinent is still *capable* of acting according to reason, i.e. capable of being rationally persuaded to change his behaviour. The incontinent, Aristotle says, is "better disposed to being persuaded (*eumetapeistos*)" (*EN* 1151a14) than the self-indulgent. Here the criterion for being 'persuadable' is the presence of regret, and the incontinent is disposed to regret (*metamelêtikos*, 1150b30) his past behaviour. Aristotle concludes from this that the incontinent is *curable* (*iatos*, 1150b32), in the rather peculiar sense that, despite the defective condition of his desiring soul, he retains the capacity of reason-responsiveness.[28] His condition is such that he retains the ability to reason correctly and deliberate about the right ends, but the judgements he forms on the basis of this ability are clouded by his strong, albeit intermittent passions. The significance of the regret that afflicts the incontinent lies in the fact that it occurs at those times when his judgement is not overshadowed by the strength of his passions; i.e. those times in which he *is* fully responsive to the guidance of right reason. This is why Aristotle compares incontinence to epilepsy, which is an intermittent condition (1150b32-35).

In *EN* I 13 Aristotle goes on to distinguish two other stages of reason-responsiveness, this time fully internalized ones:

But this [non-rational] part too seems to participate in reason (*logou metechein*), as we have said: at any rate in the continent person it is obedient (*peitharchei*) to reason – and in the moderate and

---

[27] In fact, because this reason is external in the case of the child, the child can be said to have virtue, but not to have it relative to himself (*ouk autou pros auton*) but rather relative to the fully developed being (*pros ton teleion*) and the one that has authority over him (*Pol*. I 1260a31-33).

[28] That Aristotle in these contexts has in view *rational persuasion* is clearly shown by *EN* VII 9 1151b10-16.

courageous person it is presumably still readier to listen (*euêkoôteron*); for in him it always chimes with (*homophônei*) reason. 1102b25-28[29]

In the continent, the natural disposition of the desiring part to obey and to listen to the prescribing part has reached the stage in which it *actually* obeys reason. The fact that it still needs to *obey* reason (i.e. that it needs to be told what to do), however, shows that in the continent the desiring part still has a portion of its original recalcitrant nature. Full virtue is attained when this original, recalcitrant nature of the desiring part has been fully tamed, and, as a result, the desiring part is even reader to listen to, and 'to speak with the same voice' as, reason.

Now, the soul of animals also has a desiring part, but this does not imply that it also is reason-responsive in animals. One could make a parallel mistake in the opposite direction, by assuming that, because appetite in animals does not have a share in reason, so also in human beings. These assumptions are mistaken, because the desiring part of an animal and the desiring part of a human being are not, after all, the same. They are not the same because, for Aristotle, the very nature of lower functions of the soul like perception, appetite and the emotions is infected by the highest function of reason, and partly defined in terms of this highest function. Animals, however, do not have this highest function.

But now, what about vice? Why is the *psuchê* of the fully faulty agent not mentioned by *EN* I 13 as an illustration of what is to be reason-responsive? Any account of reason-responsiveness must decide on whether reason-responsiveness is the capacity to respond to right reason (*orthos logos*), or merely to what appears as right reason; for there are good and bad fathers, and right and wrong prescriptions. That Aristotle in *EN* I 13 has right reason in view seems difficult to deny. Children, the incontinent, the continent and the virtuous all exhibit – in different degrees – responsiveness to right reason. By saying in *EN* VII that the incontinent is better disposed to being persuaded to change his mind (*eumetapeistos*) than the self-indulgent (*akolastos),* Aristotle clearly implies that the incontinent is better disposed to being persuaded *by right reason* than the self-indulgent. The self-indulgent, it is true, *has been* persuaded or convinced (*pepeistai*, 1151a13) by reason to act self-

---

[29] Rowe's translation with slight modifications.

indulgently, and like any other faulty agents, "he stands by his rational choice" (1150b30); but he is not persuaded by *right* reason. In the incontinent, on the other hand, we praise that part of the soul that has reason "because it urges (*parakalei*) him correctly (*orthôs*) and towards the best" (1102b15-16, cf. 1151a22-25, a29-b4).

Now, if reason-responsiveness is the capacity or openness of the desiring part of the soul to respond in these various degrees to *right* reason, then it is not at all clear whether the fully faulty agent has this capacity. The self-indulgent, for instance, is already rationally convinced that he must act self-indulgently; he is said to "act on the basis of his rational choice (*agetai proairoumenos*), believing that present pleasure must be always pursued" (1146b21-22); he pursues the excesses of pleasures "because of rational choice, because of themselves, and because of nothing else resulting from them" (1150a19-21); he is convinced that the reckless pursuit of pleasure is the right thing (1151a22-24) and his condition is such that he has lost the ability to form the right opinion about the ends (1151a15-19). Finally, he is not apt to regret (*ou metamelêtikos*, 1150b29). Aristotle concludes from this that the self-indulgent is not corrigible (*aniatos*, 1150a22), by which Aristotle clearly means that he cannot be *persuaded to change his mind* (i.e. he is not curable *dia to metapeisthênai*, *EN* 1146a33-34).

This claim is confirmed by *EN* X 9, where Aristotle clearly shows that the fully faulty agent cannot be persuaded to change his mind by right reason. Here (1179b4ff) the Stagirite is quite pessimistic with regard to the efficacy of *logoi* for making people decent (*pros to poiêsai epieikeis*, b5): *logoi* are impotent to stimulate 'the many' (*hoi polloi*) to excellence (1179b10), living as they do by passion (*pathei*). With regard to the vast majority of human beings, who live according to the dictates of passion, Aristotle's final verdict is categorical:

The person who lives according to passion *will not listen, nor understand*, the *logos* of the one trying to dissuade him (*ou an akouseie logou apotrepontos oud' an suneiê*); how will it be possible to persuade someone in this condition to change his mind (*metapeisai*)? *EN* 1179b26-28

Indeed, "it is not possible, or not easy, to dislodge by *logos* habits long firmly rooted in their character traits" (1179b16-18).

Now, it is true that these passages wrongly encourage us to think of vice as the mirror-image of virtue; that is, as a sort of stable character based on a fully comprehensive – though wrong – conception of the good, and the faulty agent as someone who acts *always* on the basis of such conception; as someone whose conative structure *always* cooperates with such conception, and who is *always* thinking that what he is doing is right. As some scholars have argued, this portrait of vice must be tempered by those passages in which the fully faulty agent is said to be afflicted by some sort of regret, self-reproach, and internal conflict (*MM* 1211a35-b4; *EE* 1240b12-25; *EN* IX 4 1166b7-25).[30] It is also important to note that Aristotle did not believe that the fully faulty agent is just incapable of moral reform, for he clearly believes he can be reformed by means other than *logos*, such as chastisement (1179b12).[31] Moreover, the full portrait of vice provided by these passages where the fully faulty agent is said to be afflicted by regret and at some times prone to self-reproach may even show that the fully faulty agent is far from being like Bernard Williams' callous husband, of whom Williams says that "*there is nothing in his motivational set* that gives him a reason to be nicer to his wife as things are".[32]

But whatever may be the complete portrait of the fully faulty agent one arrives at by careful consideration of these texts, such portrait must somehow remain faithful to Aristotle's unqualified verdict to the effect that the fully faulty agent, as things are, is incapable of absorbing the message conveyed by the blaming *logoi* communicated by the virtuous spectator. For all Aristotle says, children and incontinent agents can absorb this message, but not the fully faulty agent. Now, if this means that fully faulty agents are not the 'appropriate targets' of praising and blaming *logoi*, I suspect many would think that this result is unpalatable, indeed as unpalatable as it is the exclusion of children from, or the inclusion of animals and psychopaths in, the category of 'appropriate targets'. Aristotle's pessimism with regard to the reason-responsiveness of the fully faulty agent does not need to have this unpalatable result, however.

---

[30] On this see Brickhouse, T. C. (2003); Irwin, T. (1988), pp. 379-381, and (2001);
[31] This is well argued by Di Muzio, G. (2000)
[32] Williams, B. (1995), p. 39. Italics are mine.

One possibility is to understand reason-responsiveness as responsiveness to what *appears* as right reason. But this would not do: the praising and blaming *logoi* of the virtuous spectator are *right*, but do not appear so to the fully faulty agent. If so, then how can he be expected to absorb their message? The only possibility seems to be to loosen up the notion of reason-responsiveness in the direction already suggested by Aristotle. It may be that the fully faulty agent is not *actually* responsive to right reason, but in a way (if he is just faulty, and not an incurable psychopath) he retains this capacity as a second-order capacity, pretty much in the same way as I have the second-order capacity to speak German even though I have not learned it, namely, because I have the capacity to learn it. Since Aristotle does seem to accept the possibility of moral reform for the fully faulty agent, then he accepts that the fully faulty agent retains the second-order capacity for responsiveness to right reason (a capacity that animals do not have, and serious psychopaths have lost). If so, he can be counted amongst reason-responsive agents, and located at the bottom, below the reason-responsiveness of the child.

PART TWO

THE DEFEATERS OF ETHICAL ASCRIPTION

III

THE DEFINITIONS OF VIOLENCE

In this chapter I discuss in detail Aristotle's elaborate arguments for his definitions of violence (*bia*) in the *Ethica Eudemia.* The central definition of violence arrived at (what I call the 'robust' definition of violence) is unsatisfactory for many reasons, but mainly because it leaves unspecified the sort of natural impulse contrariety to which is *relevant to ethical ascription*. I offer an account of why the robust definition of violence fails in this way, and then I offer an interpretation of the contrariety condition as *contrariety to reason*. I also argue that besides this robust notion of violence, and because of some theses Aristotle endorses, he is also committed to recognize what I call a 'core' definition of violence, merely based on the externality condition. Aristotle's commitment to this core definition in turn generates two problems, to which I try to offer some solutions. One is what I call the problem of homonymy: Is 'violence' as used by Aristotle, a unified concept after all? The other is the challenge posed by Gorgias' arguments.

A. The *EE* on Violence

Proper violence or force (*bia*) may be deemed a relatively straightforward phenomenon, and consequently one that is not worth twenty or more pages of serious philosophical examination. Indeed, with the sole exception of Aristotle, no philosopher or scholar seems to have paid so much serious attention to this condition of violence (*violentum*, *force majeure*). When one is pushed off a cliff or carried away from the battle field by one's horse, one's agency plays no part in the ensuing motion, and that seems to be the end of the matter. In the hands of Aristotle, however, violence becomes an interesting topic, and one that is beset with difficulties. This is most evident in the Eudemian discussion of violence (1225a9-1225a2), which will be the main focus of this chapter.[1] This discussion is particularly interesting because it shows how Aristotle arrived at his final definition of violence.

---

[1] Aristotle sometimes uses the term '*anagkê*' ('necessity') besides '*bia*' ('violence') in the context of his discussion of violence. At 1224a13 the necessary (*to anagkaion*) and necessity (*anagkê*) make their first appearance along with the violent (*to biaion*) and violence (*bia*). In the context of this discussion, however, Aristotle shows no sign that he means '*anagkê*' to be applied to different phenomena than '*bia*', and so I will simply assume that '*anagkaion*' and '*biaion*' are here being used interchangeably, and simply talk about 'violence'.

Since action is a kind of process (*kinêsis*, *EE* 1222b29-30), it is plausible to start by treating violence in relation to processes in general. This general metaphysical picture is what seems to account for the Eudemian 'natural' approach to violence, which starts by treating violence as a property of motion and rest in general. In particular, Aristotle employs the word '*bia*' in the dative (*bia*[*i*]) which I will translate as 'through violence', to characterize primarily the manner in which motion or rest (and later, action or inaction) occur in the subject of motion,[2] namely, 'against the subject's natural tendency'. The following definition captures Aristotle's initial proposal:

V1: *X*'s motion or rest is through violence (*bia*[*i*]) iff it is against *X's* natural tendency (*hormê*).

V1 applies primarily to inanimate objects: "we say that a stone travels upwards and fire downwards through violence (*bia*[*i*]) and by necessity (*anagkazomena*), whereas when they travel according to their natural and intrinsic impulse we say that they do not move through violence" (1224a16-19).[3] We know from Aristotle's works on Natural Philosophy[4] that according to his natural system, material entities – or rather, the simple bodies of which they are constituted – have natural places and accordingly move naturally towards them unless something prevents them from doing so. And we also know that Aristotle thinks it is appropriate to apply the concept of violence (*bia*) whenever objects are thus prevented or move *from* their natural places (e.g. *Phys*. 205b5, 253b34, 255a3). The *MM* discussion of violence makes these points explicit: "For each inanimate thing there is a natural place (*oikeios topos*) assigned to them – to fire the upper region and to earth the lower. It is possible, however, to force a stone to go up and fire to go down." (1188b1-4)

---

[2] Aristotle is well aware that in such cases it is not clear whether we should speak of the 'agent' (*prattôn*) of motion or the 'patient' (*paschòn*) of motion (*EN* 1110a2; *EE* 1224a22). In order to preserve this ambiguity I will use the expression 'subject of change' or 'subject' to refer to the substance in which the enforced change occurs.

[3] I am conscious that a limitation of the term 'violence' is that it is odd to say that inanimate objects are subject to 'violence'. Another limitation is that there is no verb associated with it, as there is in the case of 'force'.

[4] See *GC.* 330b30-3; *Cael.* 276a12, 310b7; *Phys.* 253b34

It is clear that V1 is the original definition of violence: in the words just quoted from the *EE* (1224a16-19) Aristotle says nothing about the requirement that the unnatural motion of the stone or the fire be *externally caused*, as he will do subsequently in discussing animal behaviour. He makes it clear that the requirement for an external origin of motion is an addition (*proskeimenon*) to the original definition (1224b6), and this is why he describes V1 as the 'general' (*katholou*) application of the term 'violence' (1224a15).[5] But why is the requirement of an external origin unnecessary in the case of violent elemental motion? One natural (but false) suggestion is that it is unnecessary (i.e. superfluous), because inanimate objects as such *do not* have an *internal* origin of motion, in the sense that the *efficient* cause[6] of an element's natural motion towards its natural place is not internal to the element itself. This suggestion is wrong, however. The efficient cause of an element's (E) natural motion is clearly not external (at least not in the same sense in which the cause of *E*'s moving contrary to its nature - *para phusin* - is 'external'). The only external causes of *E*'s natural motion identified by Aristotle in *Phys*. VIII 4 are (i) the entity responsible for the generation of *E* and (ii) the one removing whatever may hinder *E*'s natural motion. The entity responsible for the generation of *E* is the 'external' cause of *E*'s natural motion to its natural place, but only in a very remote way, namely, in the way my French teacher is the cause of my *now exercising* the ability to speak French acquired thanks to her. And the removal of hindrances is only an 'accidental' cause of *E*'s natural motion (*Phys.* 255a24-8). The truth is that, despite Aristotle's apparent efforts to show the contrary, *E*'s natural motion seems to be fully internal to it, once *E* has been generated and there are no obstacles on its way.[7]

The requirement of an external origin does not figure in the case of *forced* elemental motion, not because all elemental motion is externally caused, but rather

---

[5] Against Kenny, who believes that V2 (see below) applies indiscriminately to all cases of violent motion (Kenny (1979), p. 38)

[6] As distinct from the formal cause (i.e. the nature of the element), which *is* internal to the element (as is clear from *Phys*. VIII 4). These two senses of 'origin' (*archê*) as formal and efficient explain why Aristotle can say that *all* natural substances (a category that includes the elementary bodies) "have *within themselves* (*en heautois*) a principle of movement and rest" (*Phys*. 192b13-14) – namely, their nature (*phusis*) – and at the same time deny that the elementary bodies are self-movers as he argues in *Phys*. VIII 4, where self-motion is said to be the prerogative of animate beings (255a6-7).

[7] As I. M. Bodnár says: "As soon as the elements are generated, and provided they are not hindered, they do not need any further causal instigation" (Bodnár, I. M. (1997), p. 102)

because – I submit – it is rendered irrelevant by a very important assumption, namely, that there is no room for 'internal' force or violence in the case of the elements (and derivatively, the inanimate entities composed of these). That is to say, there is no such thing as an *internal conflict* having a (*prima facie*) claim to internal violence in these simple entities. This is because the sheer simplicity and the continuity of their parts (*Phys.* 255a12-18)[8] prevent them from suffering any sort of internal discord or disagreement. Since there is no room for internal violence in these inanimate entities, there is no room for an internal contrariety to impulse, and thus it is *truistic* to characterize the violence exerted on elemental motion as 'external'. Note also that, because of the absence of an externality condition in V1, this definition has the consequence that if I throw a stone downwards the motion of the stone is 'natural' – as opposed to 'violent' – just as much as if I let the stone fall.

Now, when Aristotle moves on to discuss the forced movement of animals,[9] he *seems* to add a second clause, the clause about the external origin:

V2: *X*'s motion or rest is through violence iff (1) its source is in something external to *X* and (2) is against *X's* natural tendency (*hormê*).

The example given is the one of an animal that "is moved *from the outside*, contrary to its internal impulse" (1224a23), an instance of this latter being the case of a horse which is galloping straight ahead (*ep' orthon*), and is seized and diverted from its course (*MM* 1188b5-7). Later on Aristotle gives an explicit formulation of V2: "We speak of a thing being <moved> through violence (*bia*[*i*]) when it is moved or kept at rest by something external against the natural tendency (*hormê*) in the thing itself" (1224b7-9). But what is the point of adding this first clause about an external origin? Again, one may think that in the case of animals (as distinct from inanimate objects as such), the source of their

---

[8] See also *Pol*. 1254a24-b16.
[9] Aristotle begins this discussion by referring to animate things *and* animals (*epi empsuchôn kai epi tôn zô*[*i*]*ôn*, 1224a21), but I take it that the *kai* here is epexegetic. *Empsucha* is a category that includes not only animals and human beings, but also plants. Plants, however, are incapable of locomotion, which is the sort of motion Aristotle is most interested in these contexts.

voluntary motion is not external - for animals are self-movers - and consequently that externality becomes a feature *peculiar* to the violent motion of self-movers. Nonetheless, it is not at all clear – as we have seen – that the natural motion of inanimate objects (which are not self-movers) is not internally generated in the relevant way.[10] It seems wiser to think that the externality condition comes now to the fore in the case of animate beings, not because only *their* 'natural' motions are internally generated, but rather because only animals are subject to the possibility of internal conflict. This interpretation is strongly supported by the fact that the chief question Aristotle had in mind in the *EE* discussion of violence was the question about the voluntary status of the continent's and the incontinent's behaviour (1224a23-1225a2) - both of them exhibiting discord in their motivational set. Aristotle clearly thought he could solve this question by introducing the externality condition in the definition of violence (1224b11-12).

As a matter of fact, Aristotle introduces this question about the voluntariness of incontinent and continent behaviour by noticing that "in inanimate things the source <of motion> is simple, but *in animate beings it is multiple*: for desire and reason (*logos*) are not always in harmony" (1224a23-5). Clearly non-human animals do not have *logos*, but the general reference to 'animate beings' (*empsucha*) in line a24 can be explained, I think, by the fact that *some* internal conflict between impulses can be found in non-rational creatures. Just think about the dog that craves for the juicy steak on the table but does not dare to grab it for fear of being punished. Moreover, reason (in the shape of rational wish or choice) is not the only aspect of the soul that can enter into conflict with appetite (*epithumia*), the non-rational aspect par excellence. *Thumos* (anger) can also enter into conflict with appetite – recall the example of Leontius in *Republic* 439e-440a – and *thumos*, even though it was usually represented by Plato and Aristotle as being the ally of reason, and as being grounded on its evaluative judgements, is not itself an intrinsically rational impulse, and animals can also be moved by it (*EN* 1116b34; cf. *Republic* 441b). Conflict between anger and appetite, therefore, is also possible in animals.

---

[10] For discussion on this *aporia* about the source of the motion of the simple bodies, see Cooper, S. M. (1994) and Bodnár's criticisms to Cooper's view in Bodnár, I. M. (1997).

We have seen that on Aristotle's view, in inanimate things the source of motion is simple, "but *amongst animate beings* it is *multiple* (*pleonazei*) - for desire (*orexis*) and reason (*logos*) are not always in harmony" (*EE* 1224a23-25). Aristotle goes on:

So that in the case of *the other* animals what is violent is *simple* (*haploun to biaion*) – as it is in the case of inanimate things – for animals do not possess reason nor <therefore> desire that is *opposed* to it, but live by desire alone. But in the case of human beings both <reason and desire> are present, that is at a certain age, to which we attribute conduct (*to prattein*) – for we do not say that a child 'conducts himself', or a beast either; but <say this only of> someone who is *already* such as to act through reasoning. 1224a25-30

This passage must be adequately understood, if the thought that non-human animals have multiple impulses that can enter into conflict with one another is to be compatible with the present claim that, after all, the notion of violence appropriate to non-human animals is V1 (i.e. 'simple' violence). There are in fact multiple, and potentially conflicting, sources of behaviour in non-human animals, and Aristotle should not be taken as denying this obvious truth, but rather as denying – I contend – that this conflict has a *prima facie* legitimate claim to the title of 'internal violence.' We see (*horômen*, a21) that the externality condition is a separate condition of violence in the case of animals, because animals have internally conflicting impulses; the truth is that the externality condition is also irrelevant to these cases because such internal conflict does not have a *prima facie* claim to be regarded as internal violence. The question is why. What is it that renders the internal conflict between reason and non-rational desire as exemplified by the continent's and the incontinent's actions, a *prima facie* source of internal violence (i.e. one that looks as internal violence, even if this appearance is latter shown to be ungrounded), but not my dog's internal conflict between fear and appetite?

I think the *MM* example of the horse that is seized and diverted from its direct course (*MM* 1188b5-6) suggests an answer. What I have translated (following G. Stock) as '*straight* ahead´ (*ep' orthon*) can also be translated as ´in the *right* or *correct* direction'. Aristotle's idea is clearly that what forces S off a given path is 'violent' in so far as it makes sense to describe such path as the *orthos* track, but '*orthos*' has two different

connotations, one geometrical (*orthos1*) and the other one normative (*orthos2*). Applied to humans, the idea – I suggest – is that what forces a human being off a given path is 'violent' because it overcomes reason, or forces him off the *right* or *correct* (*orthos2*) track dictated by reason. It is not clear, however, that for my dog, the *right* course is to avoid the punishment, nor is it clear that the 'straight' or 'direct' (*orthos1*) course followed by the horse can be described as 'right' or 'correct' (*orthos2*) – from a non-anthropocentric perspective, that is. The fact that there are animals whose motions *can* be described as 'straight' or 'direct' (*orthos1*) explains why Aristotle *seems* to think that V2 applies to animals as well. But there are cogent reasons to think that, strictly speaking, V2 applies only to reason-responsive agents, i.e. agents who are capable of following the *right* or *correct* (*orthos2*) path. Furthermore, there is now reason to think that it is not the mere possibility of *any* internal conflict between reason and desire, but rather the possibility of conflict as we find it in the incontinent, that really has a claim to be considered violence; for only the incontinent agent (as distinct from the continent) is such as to be diverted from the right track dictated by reason. I will return to these considerations in Section E. Let us first notice another important feature of Aristotle's notion of violence.

B. Pain and the contrariety condition

V2 (the violence that, strictly speaking, applies only to human beings) says that *X*'s motion or rest occurs through violence iff (1) its source is in something external to *X* and (2) is against *X's* natural tendency (*hormê*). As I shall argue in some detail, condition (1), the externality condition, is required by the possibility of internal conflict exemplified by the incontinent, and in particular, the possibility of internal conflict *with reason* (so that the contrariety condition needs to be made more specific). Accordingly, the contrariety condition, (2), does not imply (1), the externality condition: given that human beings have two sources of motion that can enter into conflict with one another (and leaving the contrariety condition as it is, without interpretation), a given piece of behaviour can be *internally* motivated by one source and be against the other. My interest now lies in the

fact that the externality condition does not imply the contrariety condition either. The reason for this is that the contrariety condition implies what I will call 'the pain condition'. Aristotle thinks (*EE* 1224a30-31) that

(3) All behaviour brought about through violence is painful (*lupêron*).[11]

Aristotle, I think, does not mean here to suggest that the behaviour of the simple bodies is never forced – he seems to have just forgotten about them. Now, there is a clear sense in which (3) follows directly from (2), the contrariety condition, so that (3), the pain condition, is part of V2. Aristotle endorses the commonly held belief that contrariety to appetite (*to para tên epithumian*) and contrariety to temper (*to para ton thumon*) is painful (*lupêron*) – *EE* 1223a34-35, 1223b20, 1224b16-17; and he clearly thinks that contrariety to reason's prescriptions, as exhibited in the incontinent, is painful (1224b20-21). So the contrariety condition (under any possible interpretation of 'impulse') implies the pain condition (I say more about this in Chapter 7, Section B). And *vice versa*: the presence of the relevant sort of pain implies contrariety to impulse.

The significance of this, for our present purposes, is that if the relationship between pain and contrariety is one of mutual implication, as I think it is, then it is possible to determine the sort of impulse referred to in the contrariety condition through first determining the nature of the pain Aristotle has in mind. I do this in Section E, and I develop this idea further in Chapter 7. Moreover, as a result of this intimate connection between pain and the contrariety condition, one can now argue that it is perfectly possible for someone's behaviour to meet the externality condition, but not the contrariety and the pain conditions. One could imagine cases in which S's behaviour is externally manipulated and still S does not feel pain (e.g. S is glad to be carried away to Aegina by a wind), in which case – according to V2 – S would not be behaving through violence because the contrariety condition would not be satisfied (just as when I throw the stone downwards). If the contrariety condition entails the pain condition, then it is

---

[11] For the pain condition see *MM* 1188a2-3; *EE* 1223a30-35, 1224a30-31, 36-8, 1224b15-21; *EN* 1110b12; *Met*. 1015a26-33; *Rhet*. 1370a9-16.

clear that the externality condition does not entail the contrariety condition.[12] This observation will be important for my argument in Section F.

C. Continence, incontinence, and the externality problem

Now, as we are about to see, V2 is not a satisfactory definition of violence. There are two problems with V2, one relatively simple, and one serious. The relatively simple one has to do with the scope of 'external' in the externality condition, and what generates this problem is the phenomenon of continence and incontinence. Mature rational agents have two radically different internal sources of motivation, reason and non-rational desire, and these different sources of motivation "are not always in harmony" (1224a24-5).[13] Because of the existence of this discord between reason and appetite, one could object to V2 that both the incontinent and the continent act through violence because (a) "each of them acts having conflicting impulses within him" (1224a33) – the incontinent acts against his reason, and the incontinent against his appetites – and (b) each of their actions fulfils the externality condition, in so far as reason is 'external' to appetite, and appetite 'external' to reason. Aristotle's solution to this problem is quite straightforward: 'external' in V2 should not be understood as 'external to faculty X or Y', but rather as 'external to the whole soul' (b25-29):

V2*: *X*'s motion or rest is by violence iff (1) its source is in something external to X's *whole motivational set*[14] and (2) is against *X's* natural tendency (*hormê*).

---

[12] This is also suggested by the discussion of force at *Met.* V 5: "the violent (*to biaion*) is said to be what is necessary for one to do or to suffer when it is not possible to do or to suffer according to impulse, because of what is doing violence to one" (1015a34-b3), therefore implying that *it is* possible 'to suffer' according to impulse (*kata hormên*).

[13] Aristotle needs to regard both desire and reason as *impulses*. By *logos* then he may mean rational choice (*prohairesis*): Aristotle thinks that if something is to be opposed to appetite, this is indeed *prohairesis* (as shown by the case of the incontinent, *EN* 1112b13-16). By '*logos*', however, as the *MM*'s definition of violence shows (1188b6-9), Aristotle could also mean 'rational wish' (*boulêsis*), which is a sort of impulse (*hormê*, although Aristotle calls it 'desire' – *orexis*).

[14] Aristotle would say 'external to the whole soul (*holê psuchê*)', meaning by '*holê psuchê*' both rational and non-rational motivations, whether consistent or inconsistent, taken in conjunction (thus my expression 'whole motivational set').

Aristotle then makes it clear that if there were not recognizable cases of continence and incontinence, it would be superfluous to state the externality condition as an independent condition of violence, for it would be entailed by the contrariety condition (see *MM* 1188b9-12). *Contrariety to appetite* (as exemplified by the continent's behaviour), however, forces Aristotle to add an externality condition only because the contrariety condition has remained unspecified as *contrariety to impulse*, and the notion of impulse (*hormê*) encompasses both rational and non-rational inclinations. As I suggest in the next section, Aristotle uses this wide notion of impulse in V2 partly because he thinks that 'voluntary' and 'involuntary' are contradictories. Eventually, my argument will be that this wide notion of impulse poses serious problems to V2* (unnoticed by Aristotle), and these problems force him to specify the contrariety condition as *contrariety to reason.*

D. Continence, incontinence, and the pain problem

Aristotle argues that both reason and non-rational desire are natural impulses (1224b29-b37): reason is a natural source of action because it will be present if growth is allowed to continue, and non-rational desire is a natural source of action because it is immediately present from birth. These claims may suggest that, according to Aristotle, the contrariety to *natural* impulse (if we prefer to translate *hormê* thus) relevant to violence is contrariety to *both* reason *and* non-rational desire, or to *either of them*. In this section I argue that this not the case. Rather, as I will suggest in the next section, the contrariety to impulse relevant to V2* is *contrariety to reason*.

First of all, Aristotle's claim that both reason and non-rational desire are natural sources of action must be understood in the dialectical context of the *EE* II 8 discussion of violence. In particular, Aristotle is arguing against three views about the continent and the incontinent (1224a30-b2): (A) one view argues that only the continent acts through violence, because only the continent acts *against his appetite*. This view seems to assume not only a narrow, physiological notion of pain (i.e. as the thwarting of appetite), but also

that only appetite, and not reason, is a natural impulse; (B) another view[15] argues exactly the opposite: only the incontinent acts involuntarily through violence, because his appetite leads him on *without* persuading him, and persuasion (*peithô*) is the opposite of violence. This view clearly assumes a restricted notion of pain as the pain arising from *contrariety to one's rational prescription*, and also that reason is the only impulse 'natural' to rational beings as such; (C) the third view argues that both incontinent and continent persons act involuntarily through violence, because each of them acts against a natural impulse. This view then assumes a wide notion of 'natural impulse' and a wide notion of pain, both comprehending the notions assumed by (A) and (B).

Now, I think that these views are particularly problematic for Aristotle, partly because in the *EE* Aristotle holds that voluntariness and involuntariness are contradictories (*enantia*) - *EE* 1225b2, b7.[16] Their domain of application is constituted by those events that *can* be voluntary or depend on us[17] (e.g. things like our biological processes are not within our power). If *x* depends on us, and *x* is *not* voluntary, then *x* is *in*voluntary, in pretty much the way in which *x* can be said to be at rest, if *x* is the sort of thing that can be in motion and *x* is not in motion (i.e. nothing can be at rest that cannot be in motion in the first place). The problem with the aforesaid views about continence and incontinence is that they make 'voluntary' and 'involuntary' seem like mere contraries (within the specified domain). For instance, merely claiming, as Aristotle has done with regard to view (B), that the incontinent's action *x* has not an external source (being as it is *in accordance with* non-rational appetite), and assuming that it is not the result of factual error, would show that *x* is *not involuntary*; but the action would still be against THE *natural* impulse (i.e. reason, as (B) claims), thus meeting the contrariety condition. Therefore, for all Aristotle has said thus far, *x* is not voluntary either. And the same goes for the view in (A) about the continent's action, *mutatis mutandis*.

I think that Aristotle develops his wide notion of natural impulse, according to which both reason and non-rational desires are natural, precisely to avoid this

---

[15] Corresponding, it seems, to the view assumed in *EN* 1111a24-b3.

[16] I believe that Aristotle does not abandon this view in the *EN* (on this see Chapter 7).

[17] On this use general use of '*eph' hêmin*', as it occurs in the *EE* definitions of the voluntary, see Zanuzzi, I. (2007).

incongruence. If I am right, it turns out that the wide notion of natural impulse is not necessarily the relevant one for the definition of violence, but rather for the notion of *voluntary* action. Both the actions motivated by non-rational desire (like the incontinent's actions) and the ones motivated by rational desire or choice (like the continent's) are equally voluntary (1224b28, cf. *EN* 1111a24-b3), and according to nature (*kata phusin*, 1224b36).

A similar reasoning applies to the wide notion of pain, which is the counterpart of the wide notion of pleasure that Aristotle develops at 1224b16-22. Once more, the assumption that 'voluntary' and 'involuntary' are contradictories leaves room for the following objection to the views mentioned: merely claiming, as Aristotle has done concerning (B), that the incontinent's action *x* does not have an external source (being as it is *in accordance with* non-rational appetite), and assuming that it is not the result of factual error, would show that *x* is not involuntary; but this is not sufficient to show *x* is voluntary, for *x* would still be – as (B) has it – painful (i.e. the sort of pain arising from the conviction that he is doing something wrong). And the same goes for (A)'s physiological notion of pain and the continent's action, *mutatis mutandis*. If – as Aristotle wants to say – the incontinent's and the continent's behaviour is non-involuntary-but-rather-*voluntary*,[18] and if, as Aristotle often assumes (*Rhet*. I 10), pleasure is the opposite of pain and the mark of voluntariness, then there must be a suitable notion of *pleasure*, in accordance with which the incontinent person enjoys obtaining the things he craves for, and the continent the rational pleasure arising from the realization that he is in good health:

Again (*eti*) both pleasure and pain are present in the two cases: for the continent both feels pain now when acting against his appetite, but enjoys the pleasure arising from the expectation that he will be benefited later on, or <even the pleasure of being> already benefited by being in good health; and the incontinent enjoys (*chairei*) obtaining the things he craves for due to his incontinence, but feels the pain arising from expectation (*lupeitai tên ap' elpidos lupên*) thinking that he is doing something wrong (*oietai kakon prattein*). *EE* 1224b16-22

---

[18] More strictly, 'non-violent-but-voluntary' (cf. *ou bia*[*i*] *all' hekôn*, 1224b22).

Parallel then to the wide notion of natural impulse relevant to voluntariness, there is a wide notion of pleasure also relevant to voluntariness, and the resort to these two wide concepts of naturalness and pleasure is motivated by the assumption that 'voluntary' and 'involuntary' are contradictories, and the consequent need to show that the incontinent's and the continent's behaviour is non-involuntary-but-*voluntary.*

As a result of Aristotle's reply to (A), (B) and (C) supra, the Eudemian discussion on violence is brought to end leaving us with the notions of *pain* and *contrariety to natural impulse* in V2* that are simply assumed, without further argument (perhaps due to the idea that voluntariness and involuntariness are contradictories), to run parallel to the sort of *pleasure* and *concordance with natural impulse* that are the marks of *voluntariness*. In particular, the wide notion of pain:

(3)* All behaviour brought about through violence is *painful* (*lupêron*), where pain is understood in a wide sense, i.e. as including both the pain arising from contrariety to appetite and the pain arising from contrariety to reason.

But this unquestioned notion of pain, plus the idea that pain can serve as a guide for interpreting the relevant sort of contrariety, leads Aristotle into a difficult conundrum, unnoticed by him. Take the case of a diabetic person who craves for a chocolate bar (a biologically based urge) and is compelled to eat it by an external agent. Now, suppose that, on the one hand, he feels the sort of pleasure associated with the satisfaction of his non-rational appetite, and on the other he feels a 'rational' pain associated with the thought that the result is bad for him overall. The externality condition is fulfilled, because the origin of his eating the chocolate bar is external.[19] Let us grant that the diabetic agent in this example could not be temperate, because – unlike the unqualified continent – the temperate agent does not seem to care about the 'necessary' pleasures in the first

---

[19] To avoid confusion, suppose that his eating the chocolate bar is externally generated, but not a case of coercion.

place.[20] The question is whether the contrariety condition is satisfied in this case. This is what I call the 'pain problem'. In other words: is contrariety to impulse, contrariety to (i) both reason and desire taken in conjunction; or (ii) to either reason or desire; or (iii) only to reason (as I want to suggest)?

E. Solving the pain problem: moral pain and contrariety to reason.

The problem is that at this stage of Aristotle's argument for a satisfactory definition of *bia*, he cannot offer a principled answer to the pain problem. One of the reasons for this, I should like to suggest, is that (3)* represents an inadequate notion of pain, in so far as the pain in (3)* is meant to be the sort of pain relevant for involuntariness (recall that the contrariety condition can be interpreted by means of the pain condition). Let me explain this.

The argument that I develop in Chapter 7 in support of this suggestion is based on what I think is a more adequate notion of pain, as relevant for involuntariness, and it could perhaps, at this preliminary stage, be best summarized by an example. Suppose a courageous horse-rider is carried away from the battle field by his horse, which runs away terrified of the violent spectacle. He cannot come off the saddle, in which he is tightly wedged. Now, the horse-rider, we can suppose, is 'morally' distressed by the thought of losing the chance to act courageously, and by the thought that he is turning away from the battlefield *qua* an occasion calling for courageous conduct. At the same time, however, the thought that he is turning away from the battlefield *qua* the place where he could be badly wounded and even lose his life, is not distressing, but rather – we may even suppose – comforting. (These two attitudes towards the battlefield are perfectly compatible with the assumption that our horse-rider is courageous). When the question is asked to the horse-rider: 'why did you act cowardly by running away from the battlefield like a chicken?' His plea will of course appeal to the externality condition (e.g. I was tightly wedged in the saddle), but in addition to this he will – if he is indeed a courageous agent –

---

[20] By 'necessary' pleasures Aristotle means those pleasures corresponding to the satisfaction of biological desires, like the desire for food, sex, sleeping, etc. (see *EN* VII/ *EE* VI 3-4).

appeal to the *moral distress* caused by the thought that he was forsaking an occasion for courageous conduct, notwithstanding the comforting thoughts that we have also attributed to him.

The important conclusion to draw from this is that, given the sort of evaluative description of the horse-rider's behaviour B assumed by the blame, e.g. as something recognizably *cowardly*, the horse-rider's two attitudes towards the battlefield *are not equally relevant* to the involuntariness of B *thus described*. The relevant attitude for the purposes of (successfully) countering the aforesaid charge is the sort of attitude towards B *qua* cowardly, which is the horse-rider's 'moral' distress, rather than his comforting thoughts. Despite these comforting thoughts, the horse-rider's moral distress tells us that the contrariety condition is being fulfilled. Now, this is important, because the wide notion of pain in (3)* does not allow one to see this point about the primacy of *moral* or *ethically significant* pain over other sorts of pain that are irrelevant for the involuntariness of an action *recognizably faulty or virtuous*.

I take it that this sort of consideration supports the view that, contrary to what the unsatisfactory wide notion of pain appealed to in (3)* may suggest, the notion of a moral or ethically significant pain rather suggests that the contrariety condition is better understood as *contrariety to reason*:

V2**: X's motion or rest is through violence iff (1) its source is in something external to X's whole motivational set and (2) is against X'*s* rational impulse.

Now, in his Ethics Aristotle is particularly interested in cases of *perfect virtue*, and cases in which 'I was forced to do it' functions as a *plea*. These two considerations, I think, help to explain why Aristotle fails to offer the right interpretation of the contrariety condition, as well as failing to detect why the wide notion of pain in (3)* is unsatisfactory as a means to arrive at this interpretation. In the perfectly virtuous agent, like our horse-rider, practical reason and non-rational motivations are in tune. This does not have to imply that the courageous person would not, *in any way*, be distressed by the thought of

going to the battlefield, nor does it imply that the courageous person, like the continent, is such as to act according to what right reason prescribes *in spite of* his non-rational motivations to the contrary (i.e. not to do the courageous thing as such). The idea is rather that the courageous person does the courageous thing *qua* courageous (i.e. as grounding a description of it as something courageous), as reason demands, and he does it, *qua courageous*, with the full cooperation of his desiring soul. (This of course is perfectly compatible with his being afflicted by the thought of the battlefield *qua* occasion for death and physical pain). It follows from this characterization of the fully virtuous agent that, in our example of the horse-rider, where 'I was forced to do it' was understood as a plea against the charge of cowardice, the relevant contrariety to which he appeals is *ipso facto* a contrariety to his whole character (to both his rational wish *and desire*).[21] The moral pain (and the corresponding contrariety to impulse) our horse-rider feels at his being dragged away from the battle field *qua* occasion for courageous conduct, although based on his rational evaluative outlook, is felt in the whole of his character or, as Aristotle would say, *kata pasan tên psuchên* (*EN* 1166a14). If our focus is then on the perfectly virtuous agent and his plea, there is no obvious rationale for offering a particular interpretation of the contrariety condition as *contrariety to reason* (i.e. as distinct from the inclination of the non-rational part of the soul). On the contrary, the idea that 'voluntary' and 'involuntary' are contradictories contributes to his leaving the contrariety condition unspecified.

It does not follow from these facts about the perfectly virtuous person's character, however, that continent or incontinent agents, whose motivational sets exhibit a discord, cannot be sometimes subject to the sort of violence captured by V2**.[22] Consider again the case of the diabetic person who craves for a chocolate bar (a biologically based urge) and is compelled to eat it by an external agent. On the one hand, he feels the sort of pleasure associated with the satisfaction of his non-rational appetite, and on the other he

---

[21] The focus on the perfectly virtuous agent, I think, explains the contrariety appealed to in the example of the hand at 1224b14, as contrariety both to rational wish *and* appetite.

[22] *Contra* Sauvé-Meyer (1988), who thinks that the contrariety in question is contrariety to both reason and desire in harmony (i.e. moral character), thus making it impossible for robust violence to be exerted on continent and incontinent agents.

feels a 'rational' pain associated with the thought that the result is bad for him overall. Why not think that the rational pain he harbours corresponds to what I have been calling 'moral pain', in that it is, just like the pain of our horse-rider, based on an evaluative assessment of the whole situation? The pleasure and pain felt by the agent in our example exhibit his divided condition, his continent or incontinent condition. Why not think then that *this* agent's 'rational' dissatisfaction concerning the overall effects of the externally generated behaviour on his well-being, rather than the mere biological pleasure he feels from the satisfaction of his appetite, fulfils the contrariety to impulse condition, precisely in the way in which the horse-rider's moral distress does, rather than his comforting thoughts triggered by his fleeing from the battlefield *qua* source of physical pain? I will develop this argument in Chapter 7, based on what I think is the adequate notion of pain ('moral pain'). Let me now offer further reasons to think that the contrariety referred to in V2 is contrariety to reason.

First of all, recall the question raised at the end of Section A: What is it that renders the internal conflict between reason and non-rational desire in adult rational beings a *prima facie* legitimate case of internal violence, and at the same time my dog's internal conflict between fear and appetite, or anger and appetite, an illegitimate one? We saw that Aristotle's answer to this question was most plausibly grounded on the idea that what forces X *off* a given path is 'violent' in so far as it makes sense to describe such path as the *right* or *correct* (*orthos2*) track. It turns out that this is precisely what the path dictated by reason is. We saw that this was, most probably, Aristotle's reason for disregarding the internal conflict afflicting my dog, when it craves for the juicy steak on the table but does not dare to grab it for fear of being punished, as having a *prima facie* claim to internal violence; namely, that it is not at all clear that my dog's refraining from grabbing the stake for fear of punishment is 'right' or 'correct', from the dog's own point of view. Now *this* conflict between appetite's and reason's prescriptions is precisely the sort of conflict that we find in the incontinent, for it is the incontinent who, by following his appetites, acts *against his rational choice* when acting incontinently. So that, strictly speaking, it is the internal conflict exhibited by the incontinent what has a prima facie to

be regarded as internal violence (and which, accordingly, should have forced Aristotle to add an externality condition, as it is clear from *MM* 1188b9-12).

Only incontinent agents are said by Aristotle, as they were said by ordinary people and philosophers,[23] to be 'dominated' or 'controlled' (*krateisthai*, *EN/EE* 1145b36,) by something (in their case, by the 'violence' of their non-rational appetites) whereas no one has ever said of the *continent* that *he* is 'dominated' by something (i.e. by his reason).[24] The distinction between these two ethical conditions, continence (*akrateia*) and incontinence (*egkrateia*), as their Greek names suggest, is drawn in terms of *reason* controlling or *reason* being in turn controlled. The reason for this is doubtless that, in the non-rational desire/reason pair, reason has a special status, and this special status of reason is grounded on the fact that ego identifies itself with *it*. Aristotle makes this explicit at *EN* VIII:

Again, people are called continent or incontientnt (*akratês*) by reference to whether intelligence (*nous*) is in control (*kratein*) or not, which suggests that this is what each of us is. And it is actions accompanied by reason (*ta meta logou*) that people most think they have done themselves, and voluntarily. That, then, this is what each of us *is*, or this most of all, is quite clear (...). *EN* 1168b34-1169a2[25]

The fact that reason has a special status by comparison with non-rational desire shows that the 'I' in the plea 'I was forced to do it because of this or that inclination' identifies itself with reason.

One can now reconstruct Aristotle's argument in *EE* II 8 by means of the following steps: (I) Inanimate things do not exhibit internal conflict, and so the notion of violence appropriate to their motions does not require an externality condition; (II) Non-rational animals do exhibit an internal conflict, but it is not the relevant one, because it is not a conflict of their motions *with reason*, i.e. with the right or correct prescription; (III) Only incontinent agents show that the externality condition is necessary, because only the

---

[23] See also Plato's *Protagoras* 352e6-7 and *Republic* 440a.
[24] At most, Aristotle says that each of the two impulses in the continent and the incontinent is 'expelled' by the other (*ekkrouetai hup' allêllôn*, *EE* 1224b24).
[25] Rowe's translation with slight modifications. See also *EN* 116617-17 and 1178a2-4

incontinent is such as to act contrary to reason, i.e. contrary to the right or correct prescription; (IV) It follows that the final definition of violence is arrived at merely by *adding* an externality condition to what had a *prima facie* claim to be regarded as internal violence, i.e. contrariety to reason.[26]

F. The core definition of violence

To make things even more complicated, a distinction must be drawn between a robust notion of violence (the one captured by V2**), and a core notion of violence, i.e. a notion of violence merely based on the externality condition. The significance of this distinction won't be fully apparent until Chapter 7. The purpose of this section is to argue that Aristotle is in fact *committed* to recognizing a minimal or core notion of violence that does not take into account the contrariety condition at all. Suppose S's behaviour fulfils the externality condition, but not the contrariety condition (we saw in Section B that this was possible, because the externality condition does not entail the contrariety condition). So suppose the wind blows S to Aegina, and for some reason S is pleased (in the relevant sense) that this is happening. S's motion is just a motion of which S is the subject, but it is not contrary to S's reason. Is this motion voluntary? Well, of course not. Recall now that Aristotle wants to say that voluntariness and involuntariness are contradictories, *enantia* (*EE* 1225b2, b7). If *x* is the sort of thing that can be voluntary, and *x* is *not* voluntary, then x is involuntary. It follows that, if S's *externally generated motion* is *not* voluntary, and it is the sort of thing that could have been voluntarily originated (e.g. it is not a biological

---

[26] For further support of this interpretation of the contrariety conditions, see *EN* 1167b15-16; *Met*. 1015a28 and particularly *MM* 1188b6-11: The version of the contrariety condition offered by the *MM* is the disjunction "contrary to some nature *or contrary to the things he wishes* (*para phusin ti ê par' ha boulontai*)". The phrasing suggests that the disjunction first asserts the contrariety condition in a general fashion (perhaps due to Aristotle's eagerness to show that violence is a uniform phenomenon in nature), and then asserts a particular version of it (i.e. 'contrary to wish'), as an interpretation of the contrariety condition. Further, notice the preeminence of the incontinent in *MM* 1188b6-11: to accusations like 'you slept with my wife' or 'you've eaten my chocolate' the incontinent pleads 'I was forced by my appetite to act (*against my reason*)'. To maintain that this plea is ineffective and the accusations still hold, all Aristotle needs to do is to *add* an externality condition to the definition of violence, a condition that is not fulfilled by the incontinent's action. But the adding of the externality condition *does not* imply that the contrariety condition has become irrelevant or needs to be modified: it is still wish or rational choice what is standing for the self of the incontinent.

process), then it must be involuntary. Since S's motion does not fulfil the contrariety condition, however, such a motion cannot be involuntary in the sense captured by V2**. Aristotle must, therefore, recognize a core notion of violence merely based on the externality condition:

V3: X's behavior B is *involuntary* because brought about through violence iff B's source is external to X's whole motivational set.

A further reason for distinguishing a core definition of violence is the following. Aristotle illustrates the notion of violence (*bia*) with the example of someone, A, who takes hold of B's hand and hits C with it (*EE* 1224b13-14; *EN* 1135a27-8).[27] Now, when this example occurs in *EN* 1135a27-8, the reason Aristotle gives for calling the hitting of C by B's hand 'through violence' (*bia*[*i*]) is *not* that the forced movement (*M*) of the hand (which is obviously externally caused) is contrary to B's inclination: the reason he offers is that *M does not depend on* B (*ou gar ep' autô*[*i*]). This is very significant indeed, for the way Aristotle uses this notion of 'what depends on one' (*ep' autô*[*i*]) is such that whether *M* depends on S or not has nothing to do with how S happens to feel about *M*. This is evident from the role this notion of *ep' autô*[*i*] plays in Aristotle's account of deliberation – one can deliberate only about things that are *ep' autô*[*i*].[28] The idea here is that one can deliberate only about things that can be brought about through one's own agency: If you are Spartan it makes no sense for you to deliberate about the best form of government for the Scythians, because bringing the best form of Scythian government does not depend on you, i.e. it is not something you can bring about through your own agency. But how you happen to feel about the Scythians having a good government, i.e. whether you do not want this to happen or not, is absolutely irrelevant to the question of whether or not bringing about the best form of Scythian government depends on you.

Accordingly, given the tight connection I have suggested between the pain condition and the contrariety condition, it is plausible to think that Aristotle here at

---

[27] See also *Rhet.* 1369b5-6.
[28] *EE* 1226a20-33; *EN* 1112a30-31, b31-32.

1135a27-8, when he offers the example of the hand, is working with a core notion of violence, that is to say, a notion merely defined by the externality condition. Indeed, the core definition of violence is explicitly acknowledged by the *MM*:

Let this, then, be our definition of what is due to violence: those things which men are compelled to do because of something external (*ektos*) (but where the cause is internal and in themselves there is no violence). 1188b12-14

This is precisely the core definition of violence I have suggested Aristotle is committed to recognizing.[29] Now, it is worth pausing over the question of why Aristotle does not explicitly draw these conclusions in the *EE*, or in the *EN*. Once more, I think, the fact that he is particularly interested in the perfectly virtuous agent and *his* pleas, provides the clue. If 'I was forced to do it' is indeed a plea, and a plea offered by a perfectly virtuous agent S, then it follows, not only that what an apt observer is attempting to attribute to S is a faulty action, and that S did it involuntarily, but also that he did it contrary to his rational inclination, and indeed contrary to the whole of his virtuous character. If Aristotle's focus

---

[29] What about the *EN*? At 1110a1-3 Aristotle says that a piece of behaviour is *biaion* if (1) its origin (*archê*) is external, and (2) is such that 'the agent (*ho prattôn*) or rather the patient (*ho paschôn*), contributes nothing (*mêden sumballetai*) to it, e.g. "if someone is carried away by the wind or by people who have him in their power". The same claim is repeated at 1110b1-3. This is all Aristotle says in the *EN* about the violence condition per se, but it adds two points to the *EE* discussion. The first one is that it makes explicit that the external source of forced behavior need not be a human agent: one can be forced by natural phenomena like the wind. The second one is that to say that (*EN*) an agent *does not contribute in anything* to the origination of an 'action' seems different from saying that (*EE*) the practical tendencies of the (virtuous or faulty) agent is *thwarted* and *opposed* by an external force. What is the reason for this change of terminology? On what basis are we to decide upon whether the subject of motion *contributes* or not to it? On one interpretation, (2) is just a gloss on (1): to say that the subject contributes nothing is just another way of saying that the origin of the motion is entirely outside the subject. (This is Joachim's opinion: "The clearest case of compulsion is where the efficient cause is entirely outside the agent: where the agent contributes nothing to the cause." Joachim, H. (1951), p. 97. See also Irwin, T. (2007), p. 179). It is evident from Aristotle's language, however, that he believed (2) was a second condition, the non-satisfaction of which does not entail the non-satisfaction of (1). But it may also be that the connection between (2), the 'non-contribution condition' and pain provides us with an answer. Aristotle also endorses this connection in the *EN* (1110b12, 1111a25). If the non-contribution condition entails the occurrence of pain, then to say that the subject *mêden sumballetai* (does not 'join in') the motion implies that he feels pain (and conversely, to say that the subject contributes implies that he does not feel pain, correctly understood). Thus interpreted, conditions (1) and (2) correspond to V2**, the robust definition of violence. Aristotle's particular interest in this more robust definition will be explained in Chapter 7. The important point here is that, as I have emphasized, if in the Ethics Aristotle is particularly interested in cases of perfect virtue, and cases in which 'I was forced to do it' functions as a *plea*, then the agent offering this plea in these circumstances was the agent who felt the relevant pain or distress at the blameworthy event: the virtuous agent. Accordingly, I see no problem in Aristotle's failure here in *EN* III 1 to mention the core definition of violence, and accordingly no problem in assuming that the *EN* and the *EE* agree on this point.

is on the perfectly virtuous agent and his pleas, then it is easy for him to overlook the need for a core definition of violence. Nonetheless, he is committed to recognize one.

G. The problem of homonymy

Now, one could argue that it follows from the three definitions of violence that we have offered thus far, that Aristotle does not have a tolerably unified account of the nature of violence. As we have seen, he makes it clear that the requirement for an external origin of motion is an addition (*proskeimenon*) to the original definition of violence (1224b6) captured by V1, and this is why he describes V1 as the 'general' (*katholou*) application of the term 'violence' (1224a15). On my interpretation, cases of incontinence force Aristotle to develop a more complex definition of violence, as captured by V2**. And further, he is committed to recognize yet another notion of violence, one solely grounded on the externality requirement. It now looks as if 'violence' encompassed an array of disquietingly heterogeneous phenomena, or, in Aristotelian jargon, an array of phenomena which are only 'named equivocally' by the term 'violence': in consequence, a homonym.[30]

The problem of homonymy, however, can be remedied without significantly departing from the spirit of Aristotle's views on violence. The first step to take is to abandon the ambitious project – if there ever was such project – of finding a suitably general definition of violence that applies to *all* natural beings: after all, Aristotle notes that 'violence' (*bia*) is predicated of elemental motion "because of a certain resemblance" (1224b4-5), and not – we should add – strictly speaking, perhaps because when an inanimate entity moves 'naturally' and not through violence, we do not say that it moves 'voluntarily' (1224a19). Perhaps then one should concentrate one's efforts on finding a notion of violence that is the counterpart of *voluntary* motion. Focusing on voluntary motion, however, is not enough, for V1 applies to animals, and animals do move voluntarily. What Aristotle needs is rather to discard V1 as irrelevant on the grounds that

---

[30] See *Cat.* 1a1-16, on equivocal names.

it is not the counterpart of the voluntary actions of reason-responsive agents, i.e. agents who are capable of acting according to right reason. In a way, this is precisely what the *Magna Moralia* does, by assuming that the horse that is diverted from its direct course by an external agent is not subject to the robust notion of violence captured by V2**: the horse's forced motion does not deviate from the 'right' or 'correct' course. The notion of violence we are interested in is then the one that applies to agents that are capable of following the right or correct course, i.e. reason-responsive agents.[31]

But the problem of homonymy, one may still think, has not been satisfactorily remedied, because there still remains the problem of whether or not V2** and V3 are competing notions of violence. As I will argue in Chapter 7, rather than being a conmpeting definition of violence, V2** is the result of conjoining two *different* conditions blocking ethical ascription, one referring to the action (i.e. V3) and the other referring to the agent's attitudes towards his involuntary action (i.e. the pain condition, and the contrariety condition). For an elaboration of this argument, I refer the reader to Chapter 7.

Now, if I am right and V3 is indeed the core notion of violence, then we may expect Aristotle to be particularly concerned with those philosophical views which attempt to derive *all* behaviour from an external source, for this would mean, in virtue of the core notion of violence, that all behaviour is the product of violence and thus involuntary. Aristotle needs to answer this challenge if he is to succeed in offering a satisfactory account of violence, for the challenge threatens to render violence an empty concept.

H. Gorgias' challenge

The *Magna Moralia* dismisses the plea of compulsion, "I was compelled by pleasure to defile my friend's wife" (the plea of the incontinent), as "out of place" (*atopos*) on the ground that the compelled or violent is to be found 'in external things' (*en tois*

---

[31]Aristotle confines the application of V2** to those agents who are capable of *praxis* (*EE* 1224a28-30). This narrow restriction of V2** to mature rational agents, however, can be made compatible with the wider notion of reason-responsiveness: given a broad interpretation of 'contrariety to reason' in V2**, it is perfectly possible for children to fulfil the contrariety condition, in so far as children are capable of 'following the right course' dictated by an authority.

ektos, 1188b19). This is an interesting mark of the *MM* discussion: The *MM*, like the *EE*, clearly displays the belief that the clause about the external origin rules out pleas about the overwhelming force of pleasure or desire. Accordingly, Aristotle (or whoever wrote the text) is here satisfied with the argument that "when the cause is in men themselves, we will not in that case say that they are forced. Otherwise the incontinent man will have his answer ready, in denying that he is faulty. For he will say that he is forced by his appetite to perform the faulty acts" (1188b9-12). The author of the *MM* does not seem to notice, as Aristotle does in *EN* III, that such a plea can in fact be understood in terms of 'external things': "Suppose someone says that pleasant and beautiful things are violent (*biaia*), because being external they compel us (*anagkazein exô onta*)" (*EN* 1110b9-10).

My impression is that this was indeed the view put forward by Gorgias in his *Encomium of Helen*. Gorgias' purpose in this epideictic speech is to show that it is wrong to blame Helen for departing to Troy by deserting her husband and eloping with Paris. His most powerful argument is based on the idea that most of our actions are forced, because prompted by the perception of external objects. It is worth quoting Gorgias' own words:

Things that we see do not have the nature which we wish them to have but the nature which each of them happens to have; and by seeing them the mind is moulded in its character too. For instance, when the sight surveys hostile persons and a hostile array of bronze and iron for hostile armament, offensive array of the one and shields of the other, it is alarmed, and it alarms the mind, so that often people flee in panic when some danger is imminent as if it were present (§16).[32]

The same point applies to affections like pleasure, distress and motivational states: "some things naturally give distress and others pleasure to the sight. Many things create in many people love and desire of many actions and bodies" (§18).

What I call Gorgias' challenge comes from someone who argues that most of our actions, affections and motivational states (notice that Gorgias is not keen to give causal priority to the latter two over actions) are the product of violence because they can be ultimately traced back to an external origin, which Gorgias identifies with the perception

---

[32] The translation is MacDowell's.

of the relevant external object (e.g. Helen's beauty). The closest Aristotle gets to responding to Gorgias' challenge is in the following passage of *EN* III:

Suppose someone says that pleasant and fine things (*ta kala*) are violent (*biaia*), because being external they compel us (*anagkazein exô onta*). [1] If this were the case, everything would be compulsory for him: for we all do everything thanks to these things. Moreover, [2] people who act through violence and thus involuntarily, do so with pain (*lupêrôs*), whereas the ones who act because of pleasure and the good do so with pleasure (*hêdonês*). But [3] it is absurd to accuse external things (*geloion to aitiasthai ta ektos*), rather than oneself for being easily caught by these things, and to attribute the good actions to oneself and attribute the bad ones to pleasure. 1110b9-15

Of particular interest is Aristotle's second argument, (2). This argument is not good, because it does not touch the incontinent: surely the incontinent acts with pleasure in following the voice of his appetite, but he also feels rational or moral pain in disobeying the voice of his reason (and even, as Aristotle recognizes, feel the imminent pain arising from his unhealthy condition). Further, it is misleading to characterize the incontinent simply as someone who acts 'because of pleasure' (*dia to hêdu*). Hence, Aristotle's second argument seems to be addressed to someone who has the self-indulgent agent and the perfectly virtuous agent in mind. It is also clear that the first argument does not address the real difficulty raised by Gorgias' objection, although it does answer those who make use of the Gorgianic argument in an inconsistent manner, e.g. the ones who merely appeal to the external source of their blameworthy actions but want to gain credit for praiseworthy ones.

The third argument is clearly the most interesting one: it operates under the assumption that the dialectical opponent has[33] the incontinent and the self-indulgent in view, both of them acting for the sake of pleasure. Aristotle's opponent in this instance argues that these agents act involuntarily because they act for the sake of pleasure, and

---

[33] I am suggesting that Aristotle is here (1110b9-15) replying to a diversity of views. In [3] Aristotle has in mind an opponent who *attributes* (*aitiasthai*, b13) noble things (*ta kala*) to the agent and only uses the plea of externality in connection with pleasant things (*ta hêdea*). The assumed claim at the beginning of the paragraph (b9) is, however, that *both ta hêdea* and *ta kala* are forceful because external to the agent. One could of course suggest that '*ta kala*' is used with different senses in the two instances (e.g. 'beautiful things' in the first, and 'morally fine things' in the second), but this seems implausible.

111

pleasant things, i.e. the objects of non-rational desires like appetite or spirit, are ultimately the external origins of their actions. How successful is Aristotle's rebuttal? Aristotle makes two different points. One of them (the second one in the order of presentation) is to accuse his opponent of inconsistency, as in the first argument: it is inconsistent to claim that faulty agents act involuntarily because ´the pleasant compels us from the outside´ whereas virtuous ones act voluntarily, presumably because ´the fine does not compel us from the outside'; the same point about externality should apply to both. This is a very good dialectical point, for it amounts to arguing that the *onus* is on Aristotle's opponents to show why the point about externality applies to the pleasant and not to the fine. Still, the influence of the argument is far too local. The most interesting point for our purposes is the second one (and the first in the order of presentation), namely that, in D. J. Furley's words, "the proper target of blame is the deficiency of one's own character, not the external object which stimulates the action."[34] This seems to me – as it seems to Furley – to be just the right response.[35] Let us elaborate on this a bit further.

Aristotle certainly thought that a mechanical explanation of actions requires a reference to an external object. On the account of the *De Motu Animalium*, the cause of movement is "the object of pursuit or avoidance in the sphere of what is practical" (i.e. the object that can be pursued or avoided by doing something now): The object is perceived or imagined, this produces heating or cooling and the limbs, by a complicated mechanical process, move (*MA* 701a4-702a21). Aristotle does not deny that external factors figure in a causal explanation of actions. What he is denying is that we should *blame* the external stimuli *rather than* the person who reacts to them in a certain way.

Suppose that the incontinent agent offers the plea of compulsion: 'I was compelled to defile my friend's wife because perceiving her caused in me the desire to do so.' Aristotle's reply is that it is the incontinent who is blamed as 'incontinent' for having defiled his friend's wife, rather than the external stimuli, not because he could have chosen at some point not to be incontinent (as Furley thought), but rather because the

---

[34] Furley, D. J. (1977), p. 49.
[35] Recall that Furley believed that Aristotle´s claim here was based on the claim that the person in question, at some point, had in his power to acquire this or that ethical character, which would eventually determine the nature of his responses to external stimuli. See *Intro*., p. 10.

blame in question is merely a blame of him *as* 'someone who has an incontinent desire'. Reference to the external stimuli is, in this sense, *irrelevant* for the blame, in so far as it goes 'beyond character'.[36] This is what Aristotle means by saying that "we are unable (*mê echomen*) to trace our conduct back to any other sources than those within ourselves" (*EN* 1113b20-21). It is not that we simply cannot do this – indeed we can – but rather that we cannot do it if what we are engaged in doing is ethical ascription.

We can still improve on Aristotle's rejoinder that the proper target of blame is the deficiency of one's own character, and not the external object which stimulates the action. One way of doing this is by calling into question Gorgias' main assumption, namely, that whether X appears as pleasant, as beautiful, etc. to S depends solely on features intrinsic to X and extrinsic to S, that is, features of X that make no mention of S's soul or S's individual character (this is clearly one of Gorgias' own assumptions). In Aristotelian jargon, Gorgias' assumption would be that the X appears to each person the way it appears, e.g. as something pleasant, beautiful or fine, *by nature* (*phusei*). We can infer Aristotle's argument against this assumption from *EN* 1114a31-b21: the way X appears to each person is relative to that person's character (*hopoios poth' hekastos esti*, 1114a32-b1; *par' auton estin*, 1114b17): it is *par' auton* rather than *phusei*. For instance, Gorgias' claim that "when the sight surveys hostile persons and a hostile array of bronze and iron for hostile armament, offensive array of the one and shields of the other, it is alarmed, and it alarms the mind", assumes that the perception of a hostile object as alarming is like the perception of, say, a circular object. Aristotle's objection is to say that 'alarming', unlike 'circular', stands for a particular kind of relationship between the perceiver and the object: the perceiver holds an avoidance-attitude towards the object, and whether the object is alarming or not depends ultimately on the perceiver's holding this attitude towards it or not. As a consequence, an object *qua* alarming is not really an 'external thing'.

And yet, there is still a rejoinder open to Gorgias. For he can argue (as he seems to do in fact) that if one could establish that the *same* object appears alarming to *all*

---

[36] This is why Sauvé-Meyer, S. (1994) correctly argues that Aristotle's view is indeed very close to Stoic compatibilism.

perceivers (in the way Helen probably appeared beautiful to *all* men), then an object *qua* alarming is, after all, an external source. The argument perhaps would go like this: if the generalization holds, there must be a law of nature saying that certain natural properties of that object invariably cause a certain attitude towards that object under certain conditions. Even if 'alarming' stands for a relation between object and perceiver, these natural properties (e.g. the array of bronze and iron in Gorgias' example) themselves do not. If *everyone* (under 'normal' conditions) perceives these natural properties as 'alarming' (i.e. holds the attitude in question when perceiving them), whether this regularity is due to *human nature*, to the natural properties themselves, or to an interaction between these two, there may be available a suitable sense of 'external' (e.g. external to the individuals' *character*) that makes it possible to say that the object *qua* alarming constitutes an external source.

Now, if the aforementioned regularity is due to human *nature*, these perceptions may well be considered as external to the individual's *motivational set* or character broadly construed. But to reach his conclusion Gorgias needs more than this: he needs to argue, not only that (i) S's *perception* of X *qua* alarming is external to S's character, but also that (ii) S's *being alarmed* is external to his character,[37] or that (iii) S's *acting* on this perception is external in this way. The obvious rejoinder at this point is to argue that moving[38] from (i) to the more robust claims in (ii) or (iii)), is unjustified. Aristotle thinks that the courageous person, for example, can certainly *perceive*, due to his natural condition (e.g. his genetic background, etc.), a given array of shields and spears *as* alarm*ing* and fearful, and this perception may well be external to his character. But the courageous person does not *act* on this perception, nor is he *alarmed* by it.

---

[37] I say something more about (ii) in the last section of Chapter 4.
[38] As Gorgias does by saying that the hostile array of shields "*alarms the mind*, so that often people *flee* in panic when some danger is imminent as if it were present".

IV

COERCION AS JUSTIFICATION AND EXCUSE (1): THE *EE*

In this and the next chapter I develop Aristotle's theory of coercion. The main thesis I advance is that both in the *EE* and the *EN* Aristotle distinguishes between two ways in which coercion functions as a defeater of ethical ascription. In one way, coercion works as a justification, by showing that what seemed *prima facie* wrong is not so after all: I call this 'objective rational coercion'. I argue in this chapter that the *EE* fails to recognize what is distinctive about objective rational coercion as *justification*, and in the next chapter that the Nicomachean account successfully manages to bring this distinctive aspect to the fore. In a different way, the plea of coercion works as an excuse, by showing that the coerced agent, although acting wrongly, acted involuntarily: I call this 'objective overwhelming coercion'. In this and the next chapter I develop the distinctive features of this appeal to coercion, as Aristotle views it: I argue that it is not identical with the notion of psychological compulsion, that it is based on what I call the 'incommensurability' of pain and wrongness, and that it is ultimately grounded on the idea that the desire for the end of the action 'does not depend' on the overwhelmed agent.

A. Non-substantive coercion and its two varieties

In *EE* 1225a2-33 Aristotle sets out to understand an intricate group of pleas that are best, though not perhaps so neatly, captured by our pre-reflective notion of coercion. There are cases, Aristotle says, in which "people do something they believe to be painful and bad (*lupêron kai phaulon*) but are faced e.g. with flogging or imprisonment or execution if they do not do it" (1225a3-6). The *Ethica Eudemia* makes clear that the reason why Aristotle investigates the plea of coercion (*anagkê*, literally 'necessity')[1] is that

(I) people say (*phasin*) that they did these things (*tauta*) 'by necessity' (*anagkasthentes praxai*) (1225a6)

---

[1] I am conscious that a limitation of the word 'coercion' is that it seems to imply the intervention of human agency, whereas this is not the case with '*anagkê*', the term that I am now translating as 'coercion'.

By pleading coercion (*anagkê*), these people clearly imply that they acted involuntarily. But why does this appeal to *anagkê* imply or seem to imply involuntariness? The discussion of coercion follows naturally from the discussion of violence (*bia*) because the plea of coercion, "I was necessitated by the circumstances" (*anagkazomenos hupo tôn pragmatôn*, as *MM* 1188b21 puts it) *seems* to appeal to those features of robust violence that render a violent piece of behaviour involuntary - namely, externality, contrariety to impulse and the corresponding pain and reluctance.[2] This suggestion as to why cases of coercion may seem involuntary, however, is best treated as a preconception of coercion to be corrected by further philosophical scrutiny. The Eudemian analysis of coercion does this by arguing: first, that not all cases of coerced action (in what I shall call the 'non-substantive' sense) are involuntary (discussed in this section); second, amongst those that are involuntary, some (see Section B) are cases in which the agent acts under necessity and *not* through violence (*anagkazomenos kai mê bia*[*i*], *EE* 1225a17); and finally, others (Sections C-D) are involuntary due to an overwhelming inclination. Aristotle seems to think that in the latter case the language of violence (*bia*) is appropriate (1225a22, a23), and I will suggest at the end of this chapter that this is because *only these cases* bear some resemblance to violence proper, as defined in Chapter 3.

Opinion (**I**) above, that is, the opinion that "people did these things by necessity (*anagkasthentes*)" (1225a6), is challenged by another, antithetical opinion:

(**II**) They all do the action in question (*auto touto*) voluntarily. For it is possible for them not to do the action (*exesti mê poiein*), but to abide the suffering (*hupomenai to pathos*). 1225a6-8

---

[2] This motivation is reflected in the fact that, although Aristotle's preferred term in these contexts is '*anagkê*' (necessity) or '*to anagkaion*' (the necessary), he has previously defined *anagkê* at 1224b11-13 in a way parallel to his definition of *bia* (violence) given a few lines before (1224b7-8), and indeed the *MM* goes as far as to say that *to anagkaion* resides 'in the externals' (*en tois ektos*, 1188b19). It seems to me that this use of '*anagkê*' (as practically interchangeable with '*bia*') however, is mainly motivated by the dialectical nature of the preceding discussion of *bia* in *EE* 1224a13-1225a1, which shows that Aristotle is discussing with a range of philosophical and ordinary views that use the terms '*bia*' and '*anagkê*' interchangeably, as it is indeed often the case in pre-Aristotelian literature (e.g. Plato, *Cratylus* 420d-e). In my view, Aristotle uses '*bia*' and '*anagkê*' in this loose sense from 1225a2 to a17 (e.g. a2, a11, a15), and eventually he distinguishes the two concepts. Because of this I translate '*bia*' as 'compulsion' in a2-17, and then from 1225a17 to a23 I translate it as 'violence'.

Anyone well acquainted with Aristotle's dialectical method would expect Aristotle's own position to preserve what is true about each of these two antithetical opinions. Hence what follows should be naturally taken as representing Aristotle's own synthetic verdict on the controversy:

(III) And again, perhaps someone may say that some of these <actions> are <voluntary> and others are not. For [1] whatever of the things that one does without wishing them that it depends on one whether they get to be present or not,[3] these one always does voluntarily and not by compulsion. Whereas [2] those <amongst the things that one does without wishing them> that do not depend on one, one does by compulsion in a way (*bia*[*i*] *pôs*), though not without qualification (*ou mentoi g' haplôs hoti*), because one does not choose the very thing one does (*ouk auto touto prohaireitai ho prattei*), but <one chooses> that for the sake of which <one does it> (*hou heneka*). *EE* 1225a9-13

Aristotle divides 'these cases' (*toutôn*, 1225a8) in two: **(1)** There are 'cases' in which one acts voluntarily, because – as opinion **(II)** had it – one has the possibility of not doing the action (*exesti mê poiein*) and to endure the cost: it depends on one whether 'the action' is done or not (I justify the inverted commas in Section B). **(2)** And there are also cases in which 'the action' does not depend on one, that is, cases

---

[3] Kenny prefers to interpret '*hosa tôn toioutôn*' as referring to the circumstances or situation of an action, and '*huparxai ê huparxai*' (a10) as referring to the presence or absence of these circumstances. As a result, the point is that the plea of coercion is not available for the agent who voluntarily got himself into the situation in which coercion takes place (e.g. someone who voluntarily joins a criminal organization). Kenny's point is likely to generate controversy, both at the level of philosophical argument and at the level of exegesis. At the philosophical level, I think that the view attributed to Aristotle on Kenny's interpretation is not in fact Aristotle's own: the fact that A culpably joined a terrorist organization (i.e. it depended on him not to join it, and he is reasonably expected to know what joining a terrorist organization implies) does not imply that, *once* A has culpably joined the terrorist organization and is threatened with death unless he kills B, A kills B *voluntarily*. In Chapter 6 I argue that this is not Aristotle's view with regard to factual error, that is, Aristotle does not think that if S *culpably* (i.e. negligently) ignores a certain fact, and as a result S does something wrong, S does something wrong *voluntarily*. Now if this is not Aristotle's view with regard to factual error, it is natural to think that it is not his view with regard to coercion either. From the exegetical point of view, the strongest argument against Kenny's interpretation is that if Aristotle were here laying down such an important 'historical' criterion for voluntariness, there would be something deeply perplexing in his ignoring this criterion when stating his definition of the voluntary at *EE* 1225b8-10. Provided that the criterion is so essential, one would have expected Aristotle to have at least mentioned it when stating his definition of voluntariness, but no mention of this criterion is to be found (see Heinaman, R. (1988)). Finally, my rejection of Kenny's interpretation does not commit me to the implausible claim that at line a10 '*huparxai*' should be taken as meaning something like the agent doing an action. My view is that '*huparxai kai mê huparxai*' is to be taken as meaning something similar to '*ginesthai kai mê ginesthai*' at *EE* 1223a3-8 where this expression is used in a context in which Aristotle makes clear that the class of entities with regard to which the question arises whether they are voluntary or not is not necessarily restricted to actions (see also *ta huparchonta* at 1223a11-12).

in which one acts by compulsion in a way, but not without qualification.[4] (Note that Aristotle has not explicitly said that these are cases in which the agent acts involuntarily, but this follows from his claim that 'the action' does not depend on the agent, as 1225a19 shows). Aristotle explains this 'not without qualification' ('*ou haplôs*') by saying that in these cases the agent does not choose *the very thing* he *does* but rather (he chooses, *prohaireitai*) the end for the sake of which he does it.

But notice that now we need to characterize 'these cases' (*toutôn*, 1225a8) in such a way that it allows us to say that *within the same category* the distinction between (1) and (2) is to be made. This is the role of a non-substantive notion of coercion. A non-substantive definition of coercion is one that leaves it open whether people are liable to praise and blame for what they do in 'these cases', whereas a substantive notion of coercion does not leave this question open.[5] How does Aristotle characterize this non-substantive notion of coercion? Note first that he is quite careful in noticing that even those cases (within 'these cases') of coerced action that are voluntary are "such that the agent does not want to do them" (*ha mê bouletai*, 1225a10-11). In fact, Aristotle initially described cases of coercion as cases in which "people do something they *believe* (*hupolambanousin*) to be painful and bad" (1225a4-5). These are precisely the marks of the non-substantive notion of coercion: it is a notion of coercion merely based on the evaluative weighing of the alternatives *as it appears to the agent*. Of course, non-substantive coercion does not

---

[4] As noted in footnote 2, here Aristotle is using '*bia*' in a wide sense. As I suggest in Section B, Aristotle's *qualified* use of '*bia*' here ('*ou haplôs*')´ is due to the fact that he is working with the model of conditional necessity.

[5] I have in mind two substantive notions of coercion that are rejected by Aristotle. One is the normative notion. According to this notion, coercion – in the words of one of its main advocates – "always involves a definite agent (a specific individual or organization) intentionally threatening some other definite agent for the purpose of eliciting a definite action" (Fowler, M. (1982), pp. 333-4). This notion is normative because it is also required that the intentional threat be "a prima facie infringement of someone's right to shape his own life" (*ibid*, p. 332). A second notion is the compulsion notion of coercion. This is Frankfurt's view (Frankfurt, H. (1973)). Frankfurt thought that psychological compulsion was the only way to distinguish cases of coercion from ordinary cases of rational, instrumental choice. He makes it a *requirement* for coercion that the victim should not be morally responsible for submitting, and argues that a coercive threat (where the notion of 'coercive threat' covers more than the intentional threat of a human coercer) must therefore appeal to desires or motives which are beyond the victim's ability to control or which he believes to be beyond it. The normative notion is substantive in that, by definition, coercion passes the blame onto the external, intentional agent (i.e. the victim is only the instrument of the external agent's will), whereas the compulsion view is substantive in that, also by definition, coercion deprives the agent of the basic capacities required for responsibility (e.g. the capacity not to act according to one's impulses). Aristotle does not think that an external *intentional* agent is necessary for coercion, and so he rejects the normative view. And as we shall see he distinguishes coercion from compulsion, and so he rejects the compulsion view.

track Aristotle's use of *'anagkê'* in these contexts, for saying that S acted *anagkastheis* ('necessitated') clearly implies that S acted involuntarily (*akôn*), whereas non-substantive coercion does not have this implication; rather, it allows Aristotle to say that there are *cases* in which the agent sincerely *pleads* coercion, in which the acts voluntarily (i.e. he is blameworthy), and cases *within the same category* in which he acts involuntarily (i.e. he is not blameworthy).

Further light on this non-substantive notion of coercion and its two varieties is thrown by the text that immediately follows:

(**IV**) There is another distinction (*diaphora*) in these cases (*en toutois*). For if someone were to kill a man to prevent his catching him by gripping for him, it would be ridiculous for him to say that he had done something by compulsion and necessity (*bia*[*i*] *kai anagkazomenos*) – there must be <in fact> some greater evil and something more painful (*meizon kakon kai lupêroteron)* that he will suffer if he does not do it. For it will be in this way that he acts by necessity, and not through violence.[6] Or isn't it the case that he acts by nature (*phusei*), whenever he does something evil for the sake of something good, or release from a greater evil, and truly involuntarily?[7] For such acts do not depend on him. *EE* 1225a13-19

At (**IV**) Aristotle makes a further distinction between:

> (**A**) Cases in which S does X either
>
> (i) To prevent something Y that is *objectively* 'a greater evil and something more painful' (*de re*) than the one suffered by doing X or
>
> (ii) To achieve something Z (not-Y) that is *objectively* 'better and less painful' (*de re*) than not doing X

And

> (**B**), cases in which S does X either

---

[6] I follow Waltzer & Mingay in retaining the *mê bia*[*i*] in 1225a17.

[7] At a17 I have punctuated with a period after '*praxei*', and at a19 I have punctuated with a question mark after '*kai akôn ge.*' I have not accepted Russell's conjecture of '*ge*' at a18. Accordingly, I have taken '*ê ou*' at a17-18 as introducing a question (rather than taking the '*ou*' as qualifying '*phusei*', as Woods, Rackham, Solomon and Décarie do). As H. Bonitz explains, Aristotle often introduces a question with the disjunctive '*ê*' to bring a new suggestion, although in a 'modest and hesitant' way (see the *Index*, p. 313a17ff.). Bonitz offers a number of examples (*An. Pr.* 66a2; *An. Post.* 85b4, 89a16, 95b3, 98b32; *Phys.* 202b5). A good example of the question introduced by '*ê ou*' is the series of questions in the *EN* 1165b14-15, b18, b24.

(i) To prevent something Y that is *objectively* 'a lesser evil and less painful' (*de re*) than the one suffered by doing X, or

(ii) To achieve something Z that is *objectively* 'something less good and less pleasant' (*de re*) than not doing X.

Again, this distinction is possible because Aristotle has been assuming a non-substantive notion of coercion. Both (A) and (B) can be deemed cases belonging to the same category (*en toutois*, 1225a13-14) because this category is characterized by the notion of coercion sincerely employed *by the agent himself* as a plea, in so far as he sees his own situation (cf. *hupolambanousin*, 1225a4) as one in which he is being coerced by his evaluative ranking of the alternatives. According to this non-substantive notion; (a) the 'coerced action' must be *in itself* something that the agent himself considers undesirable or unwelcome (i.e. something that the agent does not want to do, 1225a10-11) but (b) where its non-performance is attached to what is, *from the point of view of the agent*, an even more undesirable result. The example in 1225a13-19 of the lad who kills someone to win the game presumably fulfils these two conditions.

Notice, however, that (a) and (b) in the previous paragraph may be deemed insufficient for the purpose of distinguishing cases of coercion in the non-substantive sense from cases of mere instrumental rationality: suppose I run out of money, and I have to fix and wash my old car in order to sell it. I can fix and wash my old car in order to sell it or not wash it and not sell it, and perhaps (a') fixing and washing my old car is in itself undesirable to me (i.e. it is burdensome, boring, time-demanding, expensive, etc.), but still (b') I may think that not selling it is even more undesirable, so I fix it and wash it. One may think then that Aristotle's non-substantive concept of coercion does not discriminate between the category of non-substantive coercion and the mere exercise of instrumental rationality (however twisted one's rational priorities may be). It is important to bear this in mind and see whether the *EN* can improve on this non-substantive notion of coercion (see Chapter 5, Section A).

Both (A) and (B) *supra*, are cases of coercion in this non-substantive sense, and the *distinction* between them is a distinction drawn in terms of *objective*

descriptions of the evils or pains involved.[8] This is very important because if praise or blame is to be bestowed on the agent pleading coercion in virtue of his weighing of the alternatives and his resulting conduct, it is going to be a praise and a blame based on objective standards, as determined by the virtuous spectator, as we have explained in Chapter 2. If what *the agent* takes to be more undesirable, more unwelcome or worse is in fact *not so* from the objective perspective of the virtuous spectator, as (B) has it, then he will be in fact blamed. If, on the other hand, what he sees as more undesirable, more unwelcome or worse *is* in fact so from this objective perspective, as (A) has it, then he won't be blamed for 'the action'.

In the next section, I will argue that the *EE* is forced to identify (1) and (2) in passage (III) – p. 117 – with (B) and (A) in passage (IV) respectively. That is, those cases of coercion (in the non-substantive sense) in which the subjective weighting of the alternatives does not correspond to the objective evaluative standard, as determined by the virtuous spectator (i.e. (B)), are in fact cases in which the agent acts voluntarily, and the action depends on him (i.e. (1)): given that blame tracks voluntariness, this agent can be blamed for his defective priorities. Those cases in which the subjective weighting is objectively right (i.e. (A)) are in fact involuntary, and the action does not depend on the agent (i.e (2)): given that blame tracks voluntariness, this agent cannot be blamed. I will call (A)-type cases, cases of *objective rational coercion* (or 'OR-coercion'): these and only these are cases of *anagkê*. I will call cases of the (B)-type, cases of *subjective* coercion. Both are cases of coercion in the non-substantive sense.

B. The *EE* on objective rational coercion

Asking whether OR-coerced *actions* are voluntarily performed or not, requires us to determine more carefully the relevant description of the action singled out for the purposes of praise and blame. Suppose a robber, after breaking into my house, wants to escape without delay from my neighbourhood and pointing a gun at my daughter, issues the threat: 'drive or I kill her'. And so I drive. Now,

---

[8] As we shall see in Chapter 5, *EN* III 1 also assumes a non-substantive notion of coercion.

'what I do' can be described in at least three relevant ways (that is, ways relevant to the question of whether the action in question is blame- or praiseworthy):

    D1   Helping the criminal escape

    D2   Saving the life of one's daughter

    D3   Helping the criminal escape in order to save the life of one's daughter

The *EE* view on what is involved in cases of OR-coercion (*anagkê*) is partly determined by the assumption that the relevant description of 'the action' is of the D1-type, corresponding to the instrumental action (what I will call, the 'non-teleological description' of the coerced action). When Aristotle introduces the Eudemian discussion on coercion by saying that in these cases "people *do something* (*ho*) they believe to be painful and bad" (1225a4), or when he states the first opinion that in these cases people "do *the action in question* (*auto touto*) voluntarily" (1225a7), or again when he says that in cases of OR-coercion "one does not choose *the very thing one does* (*auto touto ho prattei*)" (1225a12-13), he is making the following assumption:

The incomplete description assumption: What the agent *does* in cases of OR-coercion is to be described non-teleologically (e.g. D1)

    Moreover, since what the agent 'does' is captured by a non-teleological description like D1, and under this description the action is something bad (*kakon*), the *EE* also assumes the following:

The intrinsic badness assumption: In cases of OR-coercion the agent *does something bad* (*kakon prattei*, 1225a18-19; cf. *prattousi phaula*, 1228a14)

These two assumptions concerning the non-teleological description of the action and its badness, I contend, pretty much determine the *EE* conception of OR-coercion. Any plausible account of OR-coercion, I take it, has to do justice to our intuition that the objectively coerced agent should be let off the hook and should not be deemed

blameworthy. Quite the opposite, perhaps he should be deemed praiseworthy for correctly ranking the alternatives open to him. Indeed, it is reasonable to regard this intuition as a desideratum for any account of OR-coercion. Now, given the aforesaid assumptions, the *EE* can only meet this desideratum by claiming that in cases of OR-coercion, (i) "one does not choose *the very thing one does*", but (ii) "one chooses that for the sake of which one does it" (1225a12-13). Let me explain.

Aristotle makes it clear that in cases of OR-coercion, it is the instrumental action, i.e. the action in its non-teleological description (D1), that does not depend on the agent (*mê eph autô*[i], a11) and that is performed by necessity (*anagkazomenos*, a17) and thus involuntarily (*akôn*, a19). Since, as we are assuming, the instrumental action is what the agent *does*, and this is something bad, the *EE* can only conform to the aforesaid desideratum by arguing that it is nonetheless performed *involuntarily*, while at the same time allowing the agent to be credited for 'doing' something good, namely, adopting the praiseworthy end. That is, on the Eudemian view, the OR-coercee *does a bad action*, but *does it involuntarily* and thus he is not blamed for doing it, but his adoption of the end still depends on him (although Aristotle does not say this explicitly), in a way that grounds our praising him for such adoption, in so far as it reflects his (right) conception of the good. Aristotle usually explains this latter point by saying that virtue is the cause of the end of choice's being correct – and vice the cause of the end of choice's being wrong (*EE* II 12).

This is why cases of the (2)-type in passage (III) – p. 117 – correspond to cases of the (A)-type in (IV) – p. 119: If what the agent sees as more undesirable, more unwelcome or worse (e.g. letting his daughter be killed by the criminal) *is* in fact so from the objective perspective of the virtuous spectator, as (A) has it, then he should not be blamed for doing the instrumental action (e.g. helping the criminal escape). Since what he 'does' is described non-teleologically (D1), and this is something wrong, the *EE must* say – if it is to comply with our desideratum – that in cases of OR-coercion the instrumental action does not depend on the agent and that he acts 'necessitated' and involuntarily. Two questions arise concerning the *EE* conception of OR-coercion and the way in which it meets the aforementioned desideratum. First, what does Aristotle mean by saying that in cases of OR-coercion coercion the agent

'chooses' the end? And second, on what grounds can Aristotle maintain that in these cases the agent is 'necessitated' to act and acts involuntarily? Are these grounds satisfactory? Let me start with the first question.

If the *EE* position is that the OR-coerced driver does the instrumental action involuntarily, but *chooses* the end (1225a13), and "what has been chosen is voluntary" (1226b36), then the driver does the instrumental action involuntarily but achieves the end voluntarily. This in itself does not seem paradoxical (admittedly, it may sound strange if 'achieving' the end is not an action, see footnote 12). The problem arises because of the conceptual connection between S's doing *x* voluntarily and *x*'s depending on S or being 'in S's control' (*EE* 1223a4-20; *EN* 1113b20-21). To paraphrase Robert Heinaman's argument: 'if the driver *voluntarily saved* his daughter, then saving his daughter or not was in his control and he saved his daughter because he wanted to do so. Hence, he could have not saved his daughter and would have done so if he had not wanted to save her. But *how could the driver have been thus in control of whether or not he achieved the end of the action, without being in control of whether or not he performed the action itself*? How could it be that he could have stopped himself from saving his daughter when he could not stop himself from helping the criminal escape by driving the car?'[9]

What is wrong about the paradox advanced by Heinaman is that it takes '*prohaireitai*' at 1225a13 to mean 'rationally chooses, as a result of deliberation, *x qua* means to an end *y*'. If *this* is what '*prohaireitai*' means here, then of course Aristotle would be implying that it was in the driver's control whether or not *to save his daughter* (as a means to a further end): in this sense, one chooses only those *actions* that are in one's control (*EE* 1225b35-7, 1226a27-33). And contrary to what Aristotle explicitly says, he should have said that it was in the driver's control to perform the prior instrumental action by means of which he saved his daughter. Note, however, that this technical notion of *prohairesis* hasn't yet been introduced in the *EE*, and the fact that Aristotle introduces it later (*EE* II 10) in very elaborate

---

[9] Heinaman, R. (1988), p. 275..

terms strongly suggests that he can't assume that his audience is already familiar with it. '*Prohaireitai*' at a13 is then meant in the ordinary, loose sense.[10]

Now, the ordinary sense of '*prohaireitai*' perfectly allows for *the end* of the action (i.e. D2) to be the object of 'choice' *qua* end, rather than *qua* means to some further end.[11] And the end, *qua* end, is something *adopted* or *desired* by the agent.[12] This makes all the difference. As we shall see in Section D, the relevant question with regard to the 'voluntariness' of the adoption of the end is whether 'it depended on the driver' (in a sense to be specified) *not to have wanted* to save his daughter,[13] rather than the question whether he could have *saved* his daughter if he had wanted. Asking, as Heinaman does, 'how could Smith have been thus in control of whether or not *he achieved* the end of the action, without being in control of whether or not he performed the action itself?' introduces a paradox because one *achieves* the end of the action only *by doing* the action itself, whereas asking 'how could Smith have been thus in control of whether or not *he wanted* the end of the action, without being in control of whether or not he *performed* the action itself?' has a perfectly sound answer echoing the structure of hypothetical necessity: *provided* that the driver *wanted* the end of the action (and such want 'depended on him'), he was not in control of whether or not he did the instrumental action itself, which in cases of coercion is the only way to achieve the end. The fact that his adoption of the end of the action 'depends on him' (in a sense to be defined in Section D) is meant to convey the idea that it reflects his right conception of the

---

[10] Besides, it would strange to say, using this technical sense: 'S *prohairetai* that which S's action is *for the sake of*'. Although some objectives (O) are means to further ends (E), the technical sense focuses on O *as means to* E, rather than as something that some prior action is a means to.

[11] I take it that in this ordinary sense, '*prohairetai*' is used like '*hairetai*', as in *Top.*116b23ff: 'the end is regarded as more worthy of choice (*hairetôteron*) than the means to the end"; where *telos* is obviously meant *qua telos*.

[12] Although it is perhaps true that Aristotle recognized something like an action-description of type D2 above, he does not seem to recognize the *telos* (or the *hou heneka*) of an action, *qua telos*, as another *action* itself. The possibility that the agent may be *doing* two things corresponding to two *action*-descriptions: one voluntarily (e.g. saving his daughter) and the other involuntarily (e.g. helping the criminal escape), is never a problem for Aristotle (not here nor in the *EN*). In other words, when the end of the action, *qua* end, is connected to the agent's want as its object, it assumes a description in terms of a desirable characteristic, rather than as something that could be *done*; as Ackrill says, "Aristotle often gives as the object of desire (or of its species, appetite and wish) a *characteristic* (like the pleasant, the noble), and not something that could strictly be *done*" (Ackrill, J. L. (1978), p. 99). Accordingly, when Aristotle says that in the cases under consideration the agent 'chooses that for the sake of which', this does not have to imply that – on Aristotle's own standards – the agent *does* that for the sake of which *qua* such.

[13] By this of course I do not mean that the father does not feel the morally or even psychologically compelling desire to save his daughter (see Section D).

good, and the fact that the instrumental action is done involuntarily and that it does not depend on him, means that it does not reflect this conception. There is nothing paradoxical about this.

This suggestion lands us precisely on the second worry mentioned in p. 123, concerning Aristotle's grounds for saying that in cases of OR-coercion the agent does the instrumental action, which is in itself wrong, 'necessitated' and thus involuntarily. I want to suggest here that the notion of necessity (*anagkê*) employed by Aristotle in these contexts corresponds to the notion of hypothetical necessity (*anagkê ex hupotheseôs*), elaborated by Aristotle in *Phys.* II 9. This sense of 'necessity' is particularly relevant to our present discussion because, as I would like to venture, cases of OR-coercion are structurally similar to cases of hypothetical necessity. By contrast to the Democritean account of the necessity found in the natural realm, hypothetical necessity – Aristotle's own account of natural necessity – is based on a goal posited or assumed. *If*, say, a given artefact *A* is to perform a given function *F*, *A must be* constituted by material conditions *M* of type *X*. So for instance if a kitchen knife is to perform the function of cutting frozen chicken, the knife *must be* made of steel or some suitable material. The material conditions are not constitutive of the nature of the function or essence of the artefact and the goal dictated by it. Rather, given their independently established nature of *F*, the set of material conditions suitable for the performance of *F* constitutes the *only possible means* to this performance; e.g. given the function of the knife, the presence of the steel or something similar in the knife is *necessary*. Finally, the function itself is not present *because* these material conditions are present (but rather the other way around), and the mere fact that there is an object *x* which performs a certain function *F* is not necessitated by the presence of those material conditions that make the performance of the function possible. If it were, then the presence of the material conditions would not be hypothetical, but a case of 'simple' or 'unqualified' (*haplôs*) necessity, as Democritean necessity is said to be (*Phys*. 198b35): on the latter view, the material conditions operate necessarily, *on their own* so as to generate the organic structures and processes that we see.

Now, just like natural processes and artificial productions, and the organic structures and artifacts they generate, actions (*praxeis*) are goal-directed; more

precisely, they are, like natural processes and productions, goal-directed activities (*kinêseis*, *EE* 1222b29). At least this is clear. But how exactly does hypothetical necessity apply to cases of OR-coerced actions? In these latter (according to the Eudemian view) the instrumental, intrinsically wrong action *by means of which* the noble end is to be achieved or the even more shameful one avoided (e.g. helping the criminal escape) corresponds to the material conditions. What corresponds to the goal or function, on the other hand, is the end of the instrumental action (e.g. 'saving one's family'). Situations of coercion in general are such that this end can only be achieved by doing one instrumental action. But in cases of OR-coercion the end of the action (analogous to the goal or function) is not itself necessitated by the instrumental action (analogous to the material conditions). The analogy with hypothetical necessity works nicely with cases of OR-coercion, partly because these are cases in which the adoption of the end *E* can be credited to the agent and where the instrumental action is said not to depend upon the agent because (according to the analogy) *provided* that one is to attain *E*, doing the instrumental action is necessary.

This is why hypothetical necessity is not necessity without qualification (*haplôs*). This latter point is particularly important because, if the interpretation I am offering can be made to stand, it illuminates Aristotle's central claim in passage (III) – p. 117 – according to which in cases of OR-coercion 'the action' does not depend on one, and one acts "by compulsion in a way, *though not without qualification* (*ou mentoi g' haplôs*), because one does not choose the very thing one does, but that for the sake of which <one does it>" (1225a9-14). When Aristotle says that in these cases one acts by compulsion in a way and *not without qualification* (*ou haplôs*), he should be taken, I suggest, as explicitly noticing that these are cases analogous to hypothetical necessity, for hypothetical necessity is *not* absolute necessity or necessity without qualification (*ou haplôs*, *Phys.* 198b35).[14] The idea then is that, in cases of OR-coercion, the sort of necessity that we invoke in saying that the agent '*had* to do' the instrumental action (e.g. helping the criminal escape) is the sort of

---

[14] One may object to this that Aristotle is here using the word '*bia*' and not '*anagkê*'. As we have noted, however, a few lines latter Aristotle explicitly chooses to classify cases of coercion under *anagkê* and not under *bia* (1225a17).

necessity that we invoke when we say that the kitchen knife *must* be made of steel *if* it is to cut frozen chicken, where the goal of cutting frozen chicken is not itself necessitated by the material. In the same way, the father 'had to' help the criminal escape if he was going to save her daughter, but the fact that he *wanted* to save her daughter in the first place was not itself 'necessitated', and it reflects his conception of the good.

The problem that Aristotle eventually came to notice in the *EN*, I take it, is that it is not at all clear why the fact that an OR-coerced action (non-teleologically described) is 'necessitated' in this hypothetical sense renders it *involuntary*. Moreover, the Eudemian analysis suggests that cases of *subjective* coercion are cases in which the agent is to be blamed for doing the instrumental bad action (e.g. killing Smith), because it was not worth doing it given the trifling value of the end (e.g. winning the game); but the *EE* justifies this blame by claiming that in these cases the instrumental action (which in itself is wrong) depends on the agent and is *voluntary* (1225a9-11). But it is not clear why the purely subjectively coerced agent can't appeal to necessity in the sense just specified. Perhaps Aristotle wants to express the idea that the subjectively coerced agent cannot appeal to necessity, because he is not *justified.* But then the operative concept that does the work of grounding the blame in cases of purely subjective coercion is the concept of *justification*, i.e. lack of justification, rather than the concept of lack of necessity and voluntariness on its own.

As understood by John Austin in his 'A Plea for Excuses', a justification for doing *x* shows that *x* is not really wrong, while granting that *x* is voluntary. An excuse, on the other hand, grants that *x* would be the wrong thing to do, but shows that the agent did *x* is involuntarily.[15] One could say that the *EE* account of objective coercion mistakes the notion of justification for the notion of excuse.[16] As I will

---

[15] "Sometimes when I blame X for doing A, say for breaking the vase, it is a question simply or mainly of my disapproval of A, breaking the vase, which unquestionably A did; but sometimes it is, rather, a question simply or mainly of how far I think X responsible for A, which unquestionably was bad. Hence if somebody says he blames me for something, I may answer by giving a j*ustification*, so that he will cease to disapprove of what I did, or else by giving an *excuse*, so that he will cease to hold me, at least entirely and in every way, responsible for doing it." Austin, J. L. (1956-7), p. 8 (footnote 2). The underlining is mine.

[16] Michael Woods has argued that the Eudemian discussion of coercion can be read in two different ways: as a discussion on coercion as *justification* or a discussion of coercion as *excuse*. Woods, M.

suggest in the next chapter, the *EN* account of OR-coercion succeeds in doing justice to this distinction, by questioning the *EE* assumptions about the non-teleological description of 'the action' and its intrinsic badness.

What the *EE* does successfully, I think, is to elaborate a different way in which coercion can function as a plea, which does indeed correspond, I think, to the Austinian definition of excuse. To this I now turn.

C. Objectively overwhelming coercion

The suggestion I want to explore and support in this and the next section is that the *EE* also acknowledged a very different sort of reason for why, in some cases, the plea of coercion exculpates. My contention is that in *EE* 1225a17-19 (passage (IV), p. 119) Aristotle distinguishes two different cases of exculpatory coercion: (a) one in which the agent acts *by necessity* (i.e. OR-coercion) and (b) the other in which he acts 'naturally' (*phusei*). Cases of coercion in which the agent acts 'naturally' are then discussed at 1225a19ff. These are cases where the exculpating factor is the appeal to the basic nature of the agent as a human being. In an important sense, however, these are also cases of *objective* coercion (see Section D). I will call these cases, cases of 'objectively overwhelming coercion' (or 'OO-coercion').

That Aristotle is now at 1225a19ff discussing cases that do not neatly correspond to OR-coercion in the sense previously captured by '*anagkê*' is confirmed by the important notion of *to eph*' *hautô*[*i*] ('what depends on the agent') used by Aristotle to delimit these new cases of coercion, and which he defines in this context as "what the agent's nature (*phusis*) is capable of bearing" (1225a26) – this implies that what does *not* depend on the agent is, in this context, what the agent's nature is *not* capable of bearing. This notion does not fit the sense in which – as the *EE* sees it – the OR-coerced agent does the instrumental action 'by necessity' and involuntarily, and it suggests that Aristotle is now concerned with a plea of coercion that is exculpatory because the agent acted on an (objectively) *overwhelming inclination*, rather than because (as Aristotle should have said) his evaluative weighing of the

---

(1992), pp. 133-4. I think this is wrong: Aristotle distinguishes two types of coercion, one that functions as justification, and the other as excuse.

alternatives was the right one. I will discuss this in more detail below. A more immediate clue is provided by what seems to be an indispensable step in Aristotle's argument. In the process of discussing irresistible passions and beliefs, and just after having noted that "we have sympathy for them as naturally capable of constraining (*biazesthai*) nature", he says:

(**V**) A man would more seem to act through violence and involuntarily (*mallon an doxeie bia*[*i*] *kai akôn*), if he acted to escape violent rather than to escape gentle pain, and generally if to escape pain than if to get pleasure. 1225a23-25

Aristotle is here pointing at two crucial features of the coercive situation that are such as to 'generate' in the agent an (objectively) overwhelming inclination to submit to the threat. First, he is pointing at situations of *avoidance*, where what is avoided is seen as something *painful*. The idea is that situations in which the agent acts *to avoid* something *more painful*, are seen as more capable of 'generating' in the agent an overwhelming impulse to submit to the threat. This qualification seems adequate given that the dominant inclination in cases of OO-coercion is the one involved in fear (cf. *dia phobon*, *EN* 1110a4, 1116a31, 1135b5), and fear is an inclination to avoid or escape from (*hina mê*, *pheugein*) what is painful.

But this qualification is clearly not sufficient. Suppose A threatens to step on B's foot if he (viz. B) does not pull one of his own hairs (an action that will cause pain, but less pain than the prospective pain in the foot): the situation in this case is not such that it is likely to generate in B the compulsive fear of having his foot stepped on. Aristotle then reasonably notices at 1225a23-25 that the pain threatened must be *violent* (in itself) – rather than gentle, and not only more painful than the one suffered. The cases Aristotle is now interested in are then confined to cases in which the agent (i) does *x* to *prevent* some *pain y*, and (ii) where the pain threatened and to be prevented is *intense* and *bad enough* to make plausible the claim that it does not depend on the agent to endure the cost threatened.

(I say 'make plausible the claim that' instead of 'make it the case that' because we should not forget that, in the context of ethical ascription, it is the virtuous spectator who has to determine the presence of defeaters of ethical

ascription according to the nature of the case. He will of course make use of all the available evidence, including the agent's own description of the situation, but he will not have direct access to the internal, psychological mechanism (e.g. the intensity of the fear) that gave rise to the action. Thus understood, Aristotle's characterization of cases of OO-coercion is, I think, much more tenable, for rather than emphasizing the sort of features that make the coercive situation such as to *actually cause* an overwhelming impulse – thus making an easily refutable and dubious law-like generalization – it emphasizes the sort of features of the situation that make *plausible* the agent's claim that he acted on an (objectively) irresistible impulse).

Now, even though Aristotle thinks that in cases of OO-coercion the fear represents the aim of the action as one of avoidance of greater evils rather than as one of pursuit of a greater good, in cases of OR-coercion the agent can also regard the end as an object of avoidance, rather than pursuit (cf. 1225a15-20). What is essential about cases of OR-coercion or *anagkê* is not the fact that the end is represented as a positive object of pursuit, but rather that the end or goal reveals the agent's right conception of the good and his right weighing of the objective value of the alternatives left open to him. This is the basis of our condoling with him for doing an action that is blameworthy (even though the *EE* misleadingly expresses this point by saying that the instrumental action was 'involuntary'). What Aristotle says about OO-coercion in the *EE*, on the other hand, leads us to suppose that a sufficient basis for condoling with S for doing *x* (*x* being something wrong) is establishing that the *violently* painful (present, as in torture, or future) situation S tries to avoid by doing *x*, and S's corresponding fear, was *such that no one's nature could be expected to withstand* (1225a25-6). The grounds for condoling with the agent this case are quite different, for at least two reasons.

First, it seems that the pain to be avoided and the corresponding fear can be characterized as being such that 'no one's nature could be expected to withstand' *independently* of the *moral* gravity or seriousness of the instrumental action. This in turn is, I think, explained by two assumptions: (a) first, as we have seen, the end to be avoided is primarily represented by fear as something *painful* and *violent*, rather than as something *wrong* (1225a23-25). What Aristotle does not make explicit is that (b), pain and ethical wrongness may not be *commensurable*. If they are indeed not

commensurable and cannot be weighed on the same balance, the result is that one can condole with S for doing *x*, *whether 'x'* stands for revealing a trifling secret or for killing his own mother (something morally repulsive), *on the same grounds*, namely, that the pain S had to submit to (e.g. in the torture room) was such that no one's nature should be expected to withstand. Notice also that Aristotle introduces his discussion of overwhelming coercion at 1225a20 by noticing the overwhelming character of some instances of love or anger, and that these passions can provide an excuse also based on (a) and (b): If Helen did indeed something blameworthy when eloping with Paris, by leaving her nine year-old daughter behind, and if her love was indeed overwhelming, our condoling her for leaving her daughter behind involuntarily *does not* seem to depend on our first determining that eloping with Paris is *better than* staying with her daughter.

Now, it is not clear whether this is indeed the *EE* view on OO-coercion (i.e. the view that the gravity of the instrumental action is ultimately immaterial to its exculpatory character), but it does seem to me to be the most natural interpretation of it and one of its distinctive features as a plea. I will argue in Chapter 5 (Section C) that the Nicomachean account of OO-coercion can be interpreted in the same way. What I want to do next is to examine yet another aspect of OO-coercion that distinguishes it from OR-coercion.

D. Objectively overwhelming coercion and what depends on one

Since the whole question about the voluntariness of actions performed under OO-coercion turns on whether they 'depend or not on one' (1225a26), it is necessary to elucidate the meaning of this notion. When Aristotle says that *x* depends on oneself (*x* is *eph' hautô*[i]) he usually means that it depends on one to *x-or*-not-to-*x* (on this see the Appendix). One of the main obstacles to applying this notion to cases of voluntary action is that it is difficult not to analyze '*x* depends on S' as 'it depends on S to *choose* whether to bring about *x* or not'. That is to say, this bivalence seems to find its most evident expression in rational choice or in deliberation, for it is in rational choice that the agent is faced with *alternative* courses of action. This is how

Alexander of Aphrodisias understood the notion of *to eph' hautô*[i] in his *De Fato*,[17] but it is doubtless not Aristotle's own considered view on the matter (*EE* 1223a7, *EN* 1110a17-18): he clearly thinks that the notion of *to eph' hautô*[i] is applicable to voluntary actions that are not rationally chosen, which means that it is applicable to cases in which the agent is *not* deliberating whether to do *x* or not.

There is at hand another way to understand the notion of *to eph' hautô*[i], namely, in terms of subjunctive conditionals.[18] Suppose S has done *x*, and the question arises whether it depended on S to do *x* or not. Then we can say that *x* depended on S if and only if:

(**i**) If S had not wanted to do *x*, and the same conditions C1 ... C*n* had been present, S would not have done *x*.

In other words, the fact that it depends on S to do *x* or not to do x is grounded on S's capacity to determine whether *x* shall be done rather than not, merely on the basis of S's wanting to do *x* rather than not. Aristotle in *EE* II 8, however, recognized that (i) is not a sufficient condition: It may be true to say of S that if S had not wanted to do *x*, S would not – *ceteris paribus* – have done *x*, and still be true that it was not up to S *not to have wanted* to do x.[19]

That it is the notion of *eph' hautô*[i] as applied to *impulses* that becomes central to the question whether the *actions* performed *according to them* depend on one or not, is suggested by Aristotle's conclusion at 1225a30-32: "So that some thoughts and some passions do not depend on us, nor [therefore?] the actions *in accordance with* these thoughts and reasonings (*praxeis hai kata tas toiautas dianoias kai logismous*)". Indeed, this is suggested by the whole discussion of

---

[17] As he says in *De Fato* 183.27-30: "the voluntary and what depends on us are not indeed the same thing. For it is what comes about from an assent that is not enforced that is voluntary; but it is what comes about with an assent that is *in accordance with reason and judgement* that depends on us". (Sharples' translation).

[18] Following the model suggested by *Met*. IX 5 and *De An*. II 5. *De An*. says, for instance, that thinking (*noêsai*) is *ep' autô*[*i*] if one thinks whenever one *desires* to think (417b24, cf. 427b17-18).

[19] It is perhaps relevant to note that this latter condition can be read in two ways: either as 'it is not up to S *not* to want to do x' or 'it is not up to S to want *not* to do x'. It is clear that Aristotle meant the first one and not the second.

irresistible 'passions and thoughts' started at 1225a19.[20] Incidentally, notice that Aristotle here mentions 'thoughts and reasonings' along with *pathê* (see also 1225a27). I take it that this is because overwhelming impulses inevitably affect our practical deliberation or reasoning. If my fear of being tortured is indeed overwhelming (in the sense we are about to explain), then I cannot deliberate whether I should be tortured instead of, say, revealing my army's position. And here the fact that the fear is overwhelming is more basic, because it explains why I cannot deliberate, and not the other way around.

Indeed, as we have seen, in this context Aristotle defines 'what depends on one' as "what one´s nature can endure" (1225a25-26). The primary domain of application of this notion of 'what depends on one' is certainly not constituted by actions. It is worth noticing, however, that it is not *only* constituted by passions either, *pathê* like love, anger or fear - which can be said to be 'violent and beyond nature <not to have>' (a21), 'capable of constraining nature' (a22) and said not to depend on us (a31). This suggests that, fundamentally, it is (1) *the passion* or *inclination* on which the agent acts, e.g. the fear, that does not depend on him. But passage (V) – p. 130 – suggests that the domain of application of 'what depends on one' also ranges over (2) the *threatened pain* (as distinct from the fear of suffering it); that is, *the pain* is such that no one's nature could withstand. But surely (1) and (2) are connected, and in what follows the phrase 'overwhelming inclination' should be understood as reflecting this connection. Perhaps Aristotle's idea is that (1) depends on (2), that is, that we recognize an inclination like fear as 'overwhelming' or 'beyond nature' at least partly because we recognize that the pain away from which that inclination is moving us is such that no one's nature could withstand. Let us then add a second condition:

(**ii**) It depended on S not to have wanted to do *x*.

Because wanting (either desiring or rationally choosing) has its own causal antecedents, one may think that the question whether wanting *x* depends on one or

---

[20] See also *EN* 1149b3-6 (quoted in p. 136).

not is partly answered by whether or not wanting *x* is causally determined (i.e. causally determined by conditions that do not themselves depend on the agent). Aristotle, however, does not follow this route, partly because (I think) the question of whether wanting is causally determined or not is irrelevant to the question about the ethical 'ascribability' of the action based on such want. What is essential to *this* latter question is either (a) that such want proceeds from an ethical character or (b) that it is such as to be able to be inculcated in a person's character. I think that Aristotle's account of condition (ii) is devised so as to meet these requirements: it depends on S to want to do *x* or not if and only if such want is not such as to overpower human nature: if it is not "too violent for nature" (1225a22), nor "naturally capable of doing violence to nature" (1225a23). Let me explain this.

It is quite tempting to regard this notion of 'what depends on us' as 'a pain that our basic nature can bear' or 'a fear that nobody's nature should be expected to resist' as uninformative and "unfortunately obscure",[21] perhaps because it provokes the question of how to determine the degree of actual *psychological* pressure exerted by the threat that is needed for the agent thus threatened to act, as a matter of fact, by an overwhelming fear, a fear that is beyond the *individual psuchê* of the agent to resist. I don't think this question arises for cases of OO-coercion as conceived by Aristotle, however. The question that arises for Aristotle in this regard is not the question of how the virtuous spectator can determine whether the agent's fear was *in fact psychologically irresistible*. OO-coercion is *not* actual psychological compulsion – though they may overlap. This is, I think, an important and subtle distinction. The notion of an impulse (or pain) that is 'beyond nature' to resist (or withstand) is the notion of an impulse that overpowers *human nature* (*phusis*), rather than that of a desire that – as some philosophers like A. Mele maintain – overpowers *this* particular individual at a certain time.[22] It is the latter notion, not the former, that seems unfortunately obscure, but psychological compulsion is not OO-coercion.

---

[21] Irwin, T. (1980), p. 121.

[22] Consider the account of irresistible desire offered by Mele, A (1992), Ch. 5. Mele says: "When we ask whether a desire is irresistible we are asking whether *the person whose desire it is can resist it*. Moreover, since it is prima facie possible that a desire resistible by someone at one time is not resistible by him at another, the irresistibility of desires should be *relativized not only to agents but to times as well*." Mele, A. (1992), p. 86. Italics mine.

'Human nature' here functions as a normative notion. This is why I am calling this notion of overwhelming coercion *objective.* The notion of an overwhelming inclination in Aristotle's sense is such that it can be *false* to say of the overwhelmed coercee that he submitted to the threat because of a fear that was *in fact* subjectively or psychologically irresistible by him, and *still be true* to say that the fear in question was 'beyond nature to resist' in the sense captured by (a) and (b) in p. 135. That is to say, 'overwhelming' or 'beyond nature to resist' designates a property of the inclination regardless of how the agent's individual psyche was actually affected by it. Aristotle signals this by saying that in cases of OO-coercion the putative agent is pardoned because subject to an impulse that is *naturally capable* (*pephukota*) of doing violence to human nature (1225a23). Hence, there are human beings in possession of heroic virtue (*hêrôikê aretê*, *EN* 1145a20) that do not submit to the threat because they are endowed with the capacity to control the pain or distress attached to it, and yet, the fear to that pain or distress can still be described as ´naturally capable of doing violence to human nature´ because the threat is such that (a) the distress attached to it is beyond the range of pains within which virtue is expected to operate and (b) renders the example impotent for educational purposes for the very same reasons. I think that this distinction between actual psychological compulsion and Aristotelian (objectively) overwhelming impulses, does a great deal to remove the obscurity in Aristotle's notion of *ep' hautô*[*i*] as applied to pains and the corresponding impulses.

Let me insist on this distinction from a different perspective, from the perspective of *suggnomê*, sympathy. Aristotle says in the *EE* and the *EN* that cases of OO-coercion call for *suggnomê* (*EE* 1225a22-23; *EN* 1110a24). This is explained by *EN* VII (= *EE* VI), where Aristotle notes that "we pardon people more easily for following *natural desires* (*tais phusikais orexesin*), since <we pardon them more easily for following> such appetites as are more common to all human beings, and in so far as they are common (*koinai*)" (*EN* 1149b3-6). Objectively overwhelming passions are common (*koinai*) precisely in the sense that any ('good' specimen of) human being would recognize their power, and in this sense they are 'natural': they are responses that we have because of our basic nature. But the notion of human nature here is not normative in the way a mere statistical notion is: 'human nature' here means

'non-faulty, sane human nature'. The notion of human nature here is the notion used by the virtuous spectator, the one whose sympathy is authoritative. The point is then that in cases of overwhelming coercion I – as a virtuous spectator – condole with the victim not necessarily because I realize that *he* was, as a matter of fact, unable to resist, but rather because I recognize that *I* would be unable to resist it if *I*, *qua* virtuous spectator, were in his shoes.[23] Again, this objective, normative perspective of the virtuous spectator is crucial, for it rules out certain pleas of psychological compulsion that would otherwise seem successful (i.e. successful on the grounds that they render the agent's behaviour involuntary in the psychological sense of 'involuntary'). Suppose S fears to be bitten by a midge and the threat S is faced with is that unless S walks on burning charcoal S shall be locked in a room with two midges. Suppose S's terror of midges is in fact psychologically uncontrollable and overwhelming (and that S does not have some serious midge allergy). If S is led by this eccentric fear to walk on burning charcoal, it is perfectly plausible to say that S will be *blamed* for yielding to such a fear and for pursuing the wrong course of action, even though, as we have assumed, the fear in question could not in fact, as things stand, have been controlled by the agent. This is because putting up with midges 'objectively depended on him', in the sense that it is something that any non-faulty, sane human being would have done.[24] Usually, the strength of impulses and passions affecting faulty people, or the priorities they give to them (*EN* 1105b19-1106a13), are not common (*koinai*) to human beings *qua* such. The Aristotelian virtuous imputer cannot condole with S in the midge case because he (viz. the imputer) recognizes the fear as a manifestation of vice.[25] Again, psychological

---

[23] This observation is also suggested by the *Poetics*, and in particular by Aristotle's idea in *Poet.* XIII and XIV, that right tragic feeling of pity and fear can only arise when the actor is *homoios*, like ourselves (*Poet.* 1453a5, a16). This is also identified by *Rhet.* 1386a24 as a factor contributing to *heleos*.

[24] In fact, 'painful' is defined in terms of what appears as such to the virtuous agent: "But in all such matters that which appears to the good man is thought to be <as it appears to him>. And if this is correct, as it seems to be, and virtue and the good person as such are the measure of each thing, those also will be pleasures which appear so to him, and those things pleasant which he enjoys. And if the things that are tiresome for him seem pleasant to some, this is not surprising, for men are affected by many corruptions (*phtorai*) and vices (*lumai*). But these things *are not* pleasant, but only pleasant *to them* and to those in these conditions." (*EN* 1176a15-22). Ross' translation with some modifications.

[25] Feinberg has something very similar to say with regard to coercive alternatives that are only coercive because of the eccentric subjective preferences of the coercee: "Using objective standards then, we would rightly judge that B's action was voluntary enough for him to be criminally liable for it. What he is actually punished for, if it comes to that, is having the kind of character that is defined by his "cost-

compulsion, as it figures in some modern criteria for involuntariness, does not play a significant role in Aristotle's theory of coercion.

Now, an OO-impulse, that is, an impulse that any virtuous spectator (or a good human specimen) would find irresistible, is not only an impulse to do the instrumental action that is beyond human nature to resist, but also, and essentially, an impulse to adopt *the end* of the action that is beyond nature to resist adopting. That is to say, impulses that overstrain human nature are more directly connected to the end of the instrumental action, rendering the end thus adopted one that does not reflect the agent's ethical disposition, or his conception of the good. This is then another peculiar mark of OO-coercion that distinguishes it from cases of OR-oercion, where the agent's adoption of the end 'depends on him' in that it is *not* based on an impulse rooted on our basic human nature (i.e. an impulse that no one should be expected to withstand). If, instead of taking the instrumental action (i.e. *x*) as central to the notion of 'what depends on us' we take the end (i.e. *z*) as central, thus rewriting condition (ii) - p. 134 - as

(**ii***) It depended on S not to have wanted to do *z* (i.e. that for the sake of which *x* was done)

Aristotle's point would be the reasonable one that if condition (ii*) is not met – as in cases of OO-coercion – then condition (ii) is not met either, given that cases of coercion are such that the instrumental action *x* was the only way to achieve the end *z* (cf. *di' henos epiteloumenou*, *EN* 1112b17-18).

Now, granting that I am right about the distinguishing features of OO-coercion, there still remains the question: Why does Aristotle, after claiming that objective coercion does not exhibit the characteristics of violence (1225a17), persist in characterizing cases of OO-coercion as cases violence (1225a22-23)? I can only offer a tentative answer to this question, namely, 'because OO-coercion satisfies the externality condition, and thus the core definition of violence'. As we saw in the

---

orderings". Those preferences themselves are the grounds of his culpability, and he will be sentenced to imprisonment or worse, in effect, for being the kind of person who has that kind of preference hierarchy, even though, given that he *is* that kind of person, he was in an intelligible sense, *coerced*" (Feinberg, J. (1986), p. 212).

previous chapter, Aristotle understands the externality condition as 'external to the whole soul' (*EE* 1224b24-28), which we rephrased as 'external to the agent's whole motivational set'. We can now be more specific, and define a motivational set as the set conformed by all those motivations that proceed from an agent's ethical disposition (broadly understood) or that are amenable to moral education, thus excluding from this definition overwhelming desires or passions, which do violence to human nature (or, alternatively, which no human being *qua* such should be expected to resist).

V

COERCION AS JUSTIFICATION AND EXCUSE (2): THE *EN*

A. The *EN* on objective rational coercion

We are now in a position to understand the Nicomachean contribution to the theory of coercion. We saw in Chapter 4 that Aristotle in *EE* II 8 restricted cases of exculpatory coercion (i) to cases of OR-coercion, that is, cases in which the agent "does something evil for the sake of something good, or release from a greater evil"; and (ii) to cases of OO-coercion, which *EE* 1225a19ff identified as cases in which the agent acts on inclinations that are beyond nature (*huper phusin*) to resist, noting that they are cases of avoidance of particularly violent distress. *EN* III 1 opens its discussion of coercion by distinguishing these two classes of exculpatory coercion in keeping with the *EE*: cases in which one acts because of fear of greater evils (*dia phobon meizonôn kakôn*) and cases in which one acts because of something fine (*dia kalon ti*) (1110a4-5).

And just like *EE* II 8, *EN* III 1 opens its discussion of coercion with an *aporia* consisting in two antipodean opinions: "it is controversial whether these actions are involuntary or voluntary" (1110a6-7). Examples are the tyrant who, having one's family in his power, gives one orders to do something shameful or otherwise he will kill them; and the famous example of the captain who throws the goods overboard in a storm. One problem of scope immediately arises here, for – as the reader will recall - the *EE* operated with a non-substantive notion of coercion that included cases of subjective coercion (i.e. cases in which the agent's own weighing of the alternatives does not correspond to the objective perspective, being accordingly to blame for it). Aristotle seems now (1110a4-5) to be considering only cases of OR-coercion (i.e. coerced choices where there is an objectively reasonable exchange of alternatives), but this generates an inconsistency: cases of OR-coercion are always in a sense praiseworthy (i.e. the agent is praised for adopting the right end), whereas Aristotle later will say that there are *some cases of coercion* in which people are blamed because they choose to suffer something shameful for a trifling end (1111a22) – i.e. cases where there is not a reasonable exchange of alternatives (i.e.

purely subjective coercion). But we have a solution ready at hand. As in the *EE*, Aristotle is adopting the objective perspective of the virtuous spectator to describe the objective pondering of the alternatives relative to which the agent is blamed, for (genuine) cases in which the agents are blamed for making an unreasonable exchange will always be cases in which the exchange is *seen* as reasonable by the agent himself. Aristotle is better served if one reads his opening claim at 1110a4-5 as the claim that the cases now under discussion are cases in which one acts because of fear of greater evils (a claim that already allows for the fact that the evils feared might not be, as a matter of fact, greater) and cases in which one acts because of *what one believes to be* something fine.[1]

Moreover, just like the *EE*, the *EN* propounds a solution that does justice to the two parties involved in the controversy. Aristotle provides two answers, corresponding to the two categories of excusable coercion that we have distinguished. In this section and Section B I will concentrate on his departure from the Eudemian analysis of OR-coercion. I deal with his view on OO-coercion in Section C.

Aristotle affirms the part of his solution that does justice to the intuition that cases of OR-coercion are involuntary, in two ways:

A1: considered without qualification (*haplôs*), cases of coerced action are involuntary (cf. 1110a18), and considered in themselves (*kath' hauta*), cases of coerced action are involuntary (cf. 1110b3, b5).

Alternatively:

A2: "*no one* throws goods away voluntarily without qualification (*haplôs*)" (1110a9-10) and "*no one* would prefer (*an heloito*) any of these things for itself (*kath' hauto*)" (1110a19).

---

[1] Recall that the *EE* opens the discussion by describing the cases under discussion in terms of what the agents *believe* (*hupolambanousin*, 1225a4).

These claims in A2 seem rather unintelligible unless one bears in mind that Aristotle is implicitly restricting his claims to 'rational people' (*hoi noun echontes*, 1110a11). ´No one´ is here ´no one *rational*´. On the one hand, the concept of the rational person (the person having *nous*) represents the limits in the emotional make up of an agent within which ethical dispositions – whether excellent or faulty – are expected to operate (*EN* 1115b6-8, cf. 1112a19-21). In this sense, it excludes superheroes (*hêrôikê kai theia aretê*, *EN* 1145a19), bestiality (*thêriotês*, *EN* 1145a30-33) and people in pathological conditions (*EN* 1150b6-15).[2] But there is more to Aristotle's restriction of these claims in A2 to the rational person. '*Haplôs*' is always used by Aristotle to indicate the absence of any additional qualifications: *F* is said of (or belongs to) *S haplôs* if *F* is true of (or belongs to) S without adding a qualifying word or clause.[3] The qualifying phrase in this case is clearly 'on condition of (*epi*) so and so', e.g. on condition of securing the captain's own safety and the rest of the things on-board (1110a10). Surely Aristotle's idea is not simply that "no one [rational] would choose such an action, *per se* – apart from the given circumstances."[4] As it stands, I find this claim quite unintelligible, because it is not clear why 'throwing the cargo overboard' is *per se* something intrinsically irrational or undesirable,[5] as if there was an abstract class of *per se* undesirable actions. To make it intelligible, one needs to make explicit that the captain's *rational plan or interest* is to deliver the cargo: throwing the cargo overboard would be something irrational to do for the captain *relative to* the fact that the captain is a rational person *and* his rational interest or plan is to deliver the cargo.

Accordingly, Aristotle's claims in A2 are that: (1) provided the captain is a rational person, he would not throw the cargo overboard voluntarily, given his

---

[2] I do not want to translate '*hoi nun echontes*' at 1110a11 as 'reasonable' or 'sensible' people (as most translators do) because this suggests a normative, or substantive notion of rationality (i.e. the rationality of the prudent person), but I don't think this is Aristotle's point.

[3] For the use of '*haplôs*' see in particular *Top*. 115b11ff., 29ff.; *Soph. El*. 166b38ff.

[4] Joachim, H. H. (1951), p. 97.

[5] It is surprising to find interpreters simply accepting this claim as intelligible as it stands. Apart from Joachim, see also Kenny, A. (1979), pp. 31-2; and Klimchuk, D. (2002), p. 7. The only way I can make sense of these interpretations of the Nicomachean claim that cases of coercion are such that "no one would choose the action for itself", is on the basis of the Millean idea that – in Feinberg's words – "when the behaviour seems patently self-damaging and is of a sort that most calm and normal persons would not engage in, then there are strong grounds, if only of a statistical sort" for a presumption against the voluntariness of the behaviour (Feinberg, J. (1986), p. 125). Is this Aristotle's point? I doubt it. Clearly not all of Aristotle's examples of coercion are cases in which the action is *in itself* – that is, without further information about the agent – patently self-damaging and thus irrational.

rational plan, which is precisely to deliver the cargo. The point of '*haplôs*' here can be explained, I think, in the following way. The presence of a coercive factor, the presence of the storm or the threat of the tyrant, moves one from being able to have both goods at the same time - that is, delivering the cargo *and* keeping the crew safe, or not helping a criminal escape *and* preserving the life of one's daughter - to being unable to have them both at once. '*Haplôs*' here is meant to direct our attention to the pre-coercion situation in which the captain was able to choose a conjunction of the two goods: given that the captain is a rational agent and part of his rational plan in the pre-coercion situation is to deliver the cargo, he would not throw the cargo overboard in this pre-coercion situation were both goods were equally available. (2) I think Aristotle's second claim in A2 ("*no one <rational> would prefer any of these things for itself (kath' hauto)*" (1110a19)) establishes a similar point. The idea here seems to be that the captain, given that he is a reasonable person and given his initial rational plan, would not throw the cargo overboard in the *pre*-coercion situation (as opposed to "*now* and in exchange for these things (*nun de kai anti tônde*)", 1110b3-4).

We saw in the previous chapter (p. 120) that the Eudemian non-substantive notion of coercion involved the following conditions: (a) the coerced action must be *in itself* something that the coerced agent himself considers undesirable or unwelcome (*EE* 1225a4, a10-11; cf. *EN* 1110a19), but (b) where its non-performance is attached to what is, *from the point of view of the agent*, an even more undesirable result (e.g. the death of one's daughter, or loosing the game). Now, (a) and (b) were deemed insufficient for the purpose of distinguishing coerced actions (in the purely non-substantive sense) from cases of mere instrumental rationality (i.e. rationality in a non-substantive sense). If I go bankrupt, I have to fix and wash my old car in order to sell it. And perhaps (a) fixing and washing my old car is in itself undesirable to me (i.e. it is burdensome, boring, time-demanding, expensive, etc.), but still (b) I may think now that not selling it (i.e. not having the money) is more undesirable, so I fix it and wash it. Part of the Nicomachean contribution to the discussion of coercion, I suggest, consists in adding a further qualification to (a): the (non-substantive) coerced action must be in itself something that the coerced agent himself would not have done *in the pre-coercion situation, and given the agent's rational plan in that*

143

situation. Is it true that, provided I am a rational agent with a rational plan and priorities, I would *not* have fixed and washed the car voluntarily, given my rational plan before I had gone bankrupt (i.e. *haplôs*)? The relevant rational plan to which '*haplôs*' draws our attention, as we have seen, is the rational plan in the pre-coercion situation, in this case, when I had not gone bankrupt. But it is not at all clear that someone could accuse me of irrationality (in a non-substantive sense) for fixing and washing my car in the pre-bankrupcy situation and given my plan in that situation, as one could accuse the captain of irrationality if he had thrown the cargo overboard given his rational plan in the pre-storm situation (for in that situation the captain could have both kept the crew safe and deliver the cargo).[6] When I had not yet gone bankrupt, no one would have expected a rational explanation (let alone a moral justification) for fixing and washing my car!

Whatever the difference between the two claims, A1 and A2 – p. 141 – the main point is that the *EN* seems to agree with the *EE* in that 'the coerced action' when taken under its incomplete description, is *involuntary* (1110a18, b5). But this is, I take it, just a way of saying that provided the captain or the driver are rational agents, and given their rational plans, they would not throw the cargo overboard or help the criminal escape voluntarily under an incomplete, non-teleological description of these actions. The *EN*'s main contribution to the account of objective coercion is to add one crucial remainder: strictly speaking, the incomplete description of the action is not the *relevant* one for assessing whether the agent acts voluntarily or not when objectively, rationally coerced:

B: "Both 'voluntary' and 'involuntary' must be said <of an action> (*lekteon*) with reference to the time when one acts (*hote prattei*)" (*EN* 1110a15).

So that to say that the *EN* 'agrees' with the *EE* in that 'the coerced action' in its incomplete, non-teleological description (and relative to the agent's rationality) is involuntary is an overstatement, for, strictly speaking, the incomplete description of

---

[6] One may object to this proposal that it assumes the superseding of one rational plan by another, but I am here distinguishing the rational plan from the agent's fixed priorities. The presence of a coercive situation forces the agent to change his rational plan *given his fixed priorities* (the priorities do not change).

the action is *irrelevant* to the question about its voluntariness The Eudemian assumption about the non-teleological description of 'the action' is thus wrong.

There are further claims that may serve to reconstruct Aristotle's argument in B. One is that "the end (*telos*) of the action is relative to the occasion (*kata ton kairon*)" (1110a13). This claim is potentially misleading: as it stands, it may suggest that there is 'the action' on the one hand, and the end of the action on the other. Since the end is relative to the occasion, and, as Aristotle also says, "actions are in the class of particulars (*hai praxeis en tois kath' hekasta*)" (1110b7, cf. 1107a32, 1141b16), then it may seem as if 'the action' in 1110a13 refers to an action-type or universal and 'actions' in 1110b7 refer to action-tokens of that universal, thus suggesting that Aristotle's point here somehow relies on a rather mysterious and un-argued metaphysics of action. I don't think this is a feasible way to interpret Aristotle's point here.[7] The notion of occasion or opportunity (*ho kairos*, 1104a6) is not, strictly speaking, the notion of the particular circumstances "lying outside the substance of the act and yet somehow affecting it" – as Aquinas put it.[8] Aristotle's point at 1110b7 (cf. 1107a32) seems to be rather that the action *is* defined by the circumstances in which it is done (the "what, when, why, how, whom, etc." in the Aristotelian formula of the mean, 1104b22-6, 1106b21-3).[9] Since the *telos* is amongst these circumstances, indeed, since it is the *main* circumstance,[10] only the teleological description of an action – a description relative to the time of action – captures what is, strictly speaking, the action itself.

The next step in the Nicomachean argument is simply the claim that, in its teleological, proper description, the OR-coerced action is, as a matter of fact, eligible and thus voluntary. So Aristotle says that

---

[7] Taylor's translation of 1110a13 as 'the *completion* of the action is according to the occasion' suggests a metaphysical reading of Aristotle's argument, according to which the *telos* is not only the 'purpose' of the action, but also its *entelecheia*, its actuality. But this suggests that the incomplete description shows us an action *en dunamei*: does this make any sense?

[8] *Summa Theologiae*, Ia2ae. 7, 3.

[9] Indeed, it is because this occasion or moment involved in the practical considerations of the ethical agent is so particular that it is not susceptible of falling under any set of (non-empty) rules cor precepts (1104a7), which means that there is no set of (non-empty) rules or precepts in virtue of which we can define 'the substance' of the act.

[10] I further elaborate on this claim in Chapter 6, Section B.

C: Under their teleological, proper descriptions, these actions are voluntary, "since (*gar*) they are eligible (*hairetai eisin*) at the time when they are done (*tote hote prattontai*)" (1110a12-13)

By claiming that "these actions are *eligible* (*hairetai eisin*) at the time when they are done (*tote hote prattontai*)" Aristotle may well be addressing the Eudemian claim that in cases of OR-coercion the agent "does not choose (*ouk prohaireitai*) the very thing he does" (1225a13): it may be that 'the very thing he does', when this is separated from the actual end, is not something eligible (*haireton*), but given that (i) as the *EE* itself admits, these are cases in which the agent himself *chooses* to adopt the end of the action (1225a14) and (ii) a teleological description of the action is the only relevant description, 'the very thing he does' (i.e. doing-*x*-for-the-sake-of-*z*) turns out to be voluntary, because eligible in the OR-coercive situation.[11]

The best way to understand Aristotle's final claim in C is, I think, in the light of the desideratum mentioned in Chapter 4 (Section B), according to which the OR-coerced agent should be let off the hook and should not be deemed blameworthy. On the contrary, as we saw, he should be deemed praiseworthy for correctly ranking the alternatives open to him. We saw in Chapter 4 that since the *EE* assumes that what the agent 'does' is to be captured by a non-teleological description, and that under this description 'the action' is wrong, the only way the *EE* can accommodate

---

[11] All commentators would disagree with me here, for it has been traditionally thought that the main *EN* argument for the claim that OR-coerced actions are voluntary is that the source of movement of the instrumental parts of the body is in the agent, and it is therefore 'up to' the agent to act or refrain from acting (1110a15-18). But this traditional interpretation fails to acknowledge claims like the one at 1110a12-13 and 1110b4. I think the argument about the instrumental limbs is a second argument. Incidentally, this second argument from the voluntary movement of the limbs is seriously flawed. The argument is this:

  (i) If the source of x is in S himself, it depends on S to do or not to do x

  (ii) (In the cases under discussion) the source of S's moving the instrumental parts of his body is in S himself (*en autô*[*i*])

  (iii) (In the cases under discussion) it depends on S to move or not to move the instrumental parts of his body.

     It is clear that with this argument Aristotle intended to support a much stronger conclusion than (iii), namely (iii*) that it depends on S e.g. *to jettison the cargo or not to jettison the cargo* – and not only to set in motion the instrumental parts of his own body *by means of which* he jettisoned the cargo. That (iii) does not entail this stronger conclusion is pretty obvious: I can jettison the cargo in the mistaken belief that I am (not jettisoning the cargo but) getting rid of rubbish (suppose I am unaware of the storm), in which case, although I am in control of my bodily motions, I am not jettisoning the cargo voluntarily. That is to say, the control of my bodily motions only serves to support the voluntariness of my jettisoning the cargo if I move my body *so as to bring* [*i.e. with the intention of bringing*] *it about that the cargo is jettisoned.*

our desideratum is by arguing that what the OR-coercee 'does' is involuntary. We saw that this was an unsatisfactory account of the appeal to OR- coercion as a plea, because Aristotle did not offer convincing grounds for thinking that the OR-coerced agent acted *involuntarily*, and he failed to capture what seemed central to OR-coercion, which is its *justificatory* role. Incidentally, working out what the role of ethical justification (as different from excuse) is, and characterizing cases of OR-coercion in terms of it, may now seem like an obvious move to make, but apparently, at the dawn of philosophical analysis it required a great deal of conceptual work and a high level of abstraction. This conceptual work is carried out in the *EN*, which remedies the deficiency in the *EE* account by arguing that the latter is wrong to assume that what the coerced agent 'does' is to be captured by a non-teleological description of the action, and by claiming that under the relevant, teleological description – as we shall see in the next section – the action is *not* wrong (against the Eudemian assumption of the intrinsic badness of the coerced action). This is why the *EN* is in a position to do justice to our desideratum by claiming that the OR-coerced agent acts voluntarily. By doing this, the *EN* account amounts to a recognition of OR-coercion as justification, as distinct from excuse.

Aristotle's way of developing this concept of justification has its peculiarities, however. One justifies oneself (in the Austinian sense) whenever one admits that one has acted voluntarily, but shows that what one 'did' was not, 'after all', wrong. But there is more than just one way to account for what is going on in justification, thus described. For instance, someone may argue that (1) the driver *does two things* (as in the *EE*): (i) helping the criminal escape and (ii) saving his daughter; and that (2) he does them *voluntarily* (as in the *EN*), but that (3) given the objective value of the alternatives, the badness of the first is 'outweighed' by the goodness of the second and accordingly the agent is praised only *for the latter*. The *EN* notion of justification is not this: objective coercion justifies the agent because, although the agent acts voluntarily, *he does nothing wrong voluntarily* (remember: the driver does not perform an action whose full description is 'helping the robber escape'). On the contrary, given his objectively right pondering of the alternatives, *what he does* (i.e. *the whole* of what he *does*) is *praiseworthy*. In the next section I attempt to offer firmer support to these claims.

B. The principle of preference

This is the occasion to turn to the Aristotle's acknowledgment of what I will call 'the principle of preference' (PP). The principle is stated in the *Ars Rhetorica*:

all that men do voluntarily will be either that which is or seems good, or that which is or seems pleasant. For I reckon among good things (*en tois agathois*) the removal (*apallagê*) of that which is evil or seems evil, or the exchange of a greater evil for a less (*anti meizonos kakou, elattonos metalêpsin*), because these two things are in a way preferable (*haireta*); in like manner, I reckon among pleasant things (*en tois hêdesin*) the removal of that which is or appears painful, and the exchange of a greater pain for a less. *Rhet*. 1369b23-28

I would like to venture that since Aristotle in the *Ethica Nicomachea* explicitly adopted the complete description of 'the action', he came to endorse PP. The principle has two parts: (1) The *removal* of that which *is* evil/painful or seems evil/painful, which *is* good/pleasant because preferable; and (2) the exchange of a greater evil/pain for a less (*metalêpsis x anti y*), which again *is* good/pleasant because preferable. Notice that Aristotle also formulates PP in terms of *apparent* evils or pains, but since our present topic is objective coercion, we could leave this aspect of the principle aside.

The passage in the *Ars Rhetorica* provides us with some hints to the effect that Aristotle considers PP to be applicable to cases of objective coercion. In the *EE*, for instance, he describes cases of OR-coercion as cases in which one does something evil for the sake of the 'removal' of or better, the 'distancing oneself from' (*apolusis*, a synonym of *apallagê*) something more evil or painful (*EE* 1225a18-19; a16), e.g. the driver helps the criminal escape in order to *prevent* his daughter from being killed. As to (2), the language here is strongly reminiscent of the *EN* III 1 discussion of coercion, where Aristotle says that OR-coerced actions "*now and in exchange for these things are eligible* (*anti tônde haireta*)" and "*now and in exchange for these things* (*anti tônde*) are voluntary" (1110b3-6, a30, a19-23). This is precisely what Aristotle is saying in (2): first, it is the *exchange* (*x anti y*) of greater evils (*meizonôn kakôn*, 1110a4, cf. *EE* 1225a16) for lesser ones that Aristotle has in

mind, and more importantly, it is *the exchange itself* ('helping a criminal escape *anti* letting one's family die') what he considers as eligible (*haireton*), and therefore as good; this latter being precisely the reason offered in the *Ars Rhetorica* to consider the exchange in question as a good one.

Notice also that Aristotle is careful to formulate PP both in terms of 'bad'/'good' and 'painful'/'pleasant', just as he does in his discussion of OR-coercion, both in the *EE* and the *EN*. From the point of view of the Ethics, both characterizations can equally ground praising or blaming *logoi*: for instance, if what the agent sees as more painful is in fact so from an objective perspective, he is as praiseworthy as he would be if he had seen it as morally worse. This is why I haven't made much of the difference between these different ways of valuing the alternatives open to the OR-coerced agent. What is interesting about the formulation of PP in these two sorts of evaluative terms, is that Aristotle does not mention the possibility that 'bad' and 'painful' could be weighed on the same balance. Notice that this is compatible with (and it rather tends to support) our assumption in Chapter 4 (Section C) that pain/pleasure and badness/goodness are not commensurable.[12]

PP implies that what the OR-coerced agent does (i.e. the exchange itself) is *good*. That the *EN* adopts PP is also clear from the following considerations concerning shame. Notice first that, even though Aristotle describes the person threatened by the tyrant as *doing something shameful* (*aischron ti praxai*, 1110a5), he is logically committed to the view that, strictly speaking, saying that he 'does' something shameful is merely representing the undesirable and incomplete, non-teleological aspect of the 'coerced action', but the *EN* is particularly insistent upon the fact that paying attention only to this aspect is wrong. But then why saying that the victim of the tyrant's threat *does* something shameful? Is Aristotle unaware of the logical consequences of his view in the *EN*?

That he is not unaware of these logical consequences, can be shown in two alternative ways. First, characterizing the instrumental action as something

---

[12] This question about incommensurability is closely connected to the problem of *akrasia*, as Plato (in the *Protagoras*) and Aristotle understand it. On this see Wiggins, D. (1978/79) and Charles, D. (1980), Chapter 3.

'shameful' (*aischron*) seems to have different implications to characterizing it as something 'bad' (*kakon*, *phaulon* – as the *EE* does: 1225a18-19, 1228a14). One important consequence of expressing the negative aspect of the coerced action in terms of shame derives from the fact that 'shame dwells in the eyes' (*Rhet.* 1384a18).[13] Aristotle says that "shame is an *appearance* about dishonour (*peri adoxias phantasia*, *Rhet.* 1384a13)", and by this he means that being ashamed of *x* is grounded on a regard for people's opinion (*doxa*) about *x* – in particular the opinion of people whom one esteems (1384a14) – and the belief that *x* will be, is, or was ill-regarded by these people. Accordingly, one possibility is that by saying that the victim of the tyrant's threat *does* something shameful, Aristotle is merely attributing to him the reasonable belief that people whom he esteems will regard his action badly (and perhaps one should add: 'provided they don't have access to the full picture of S's situation'). This is perfectly compatible with S's action being in fact something good, as Aristotle seems to admit in *EN* 1128b23-5.[14]

I must confess that I am not sure whether this is in fact Aristotle's view, however. For Aristotle also seems to think, at least in the *Ars Rhetorica*, that we are only ashamed of doing *x* when, besides seeming to bring disrepute and disgrace amongst highly esteemed people, *x* displays *in fact* (as opposed to merely appearing to display) a character flaw (see the list at *Rhet.* 1383b-1384a). So perhaps the key consideration is that, even if 'shameful' was thought to be incompatible with 'objectively good', there is reason to think that, according to the *EN* the 'shameful thing' is not something that the coerced agent *does*, but rather something that the coerced *endures* (*hupomenein*, *EN* 1110a21, a22, a30). As Nielsen has pointed out, the verb "*hupomenein* has unmistakable connotations of passivity. In performing a coerced act, we choose to *endure* something shameful or painful for the sake of the fine or in order to avoid a greater evil. We do not, to Aristotle's mind, voluntarily *perform* an action that is shameful under the circumstances".[15] If this is right, when

---

[13] The numbers here refer to Ross' edition.

[14] Plato for instance, admits that one can do something shameful voluntarily, but not something wrong voluntarily (*Protagoras* 345d)

[15] Nielsen, K. (2007), p. 283. Nielsen is arguing against Stocker's and Nussbaum's view that Aristotle in the *EN* allows for dirty hands cases, that is, that he allows for the justification of actions that *are in fact* morally repugnant and always wrong and even recognized as such by the hypothetical agent. I think his arguments against this view are very cogent.

Aristotle says that the tyrant's victim *does* something shameful, he should be taken as meaning that he *endures shame* by doing what is, in fact, a good action.

I think I am entitled to conclude, then, that the *EN* successfully isolates the distinguishing feature of objective coercion as a plea, namely, its justificatory character. It does so in its own, peculiar way, though. The idea is not that the objectively coerced agent is justified in *doing something wrong* voluntarily: he does not voluntarily do something wrong, but something good, given the teleological description of what he does, and the adoption of PP.

C. Objectively overwhelming coercion and matricide

In this section I want to discuss the Nicomachean view on objectively overwhelming coercion (OO-coercion). After having observed that in some cases of coercion (in the non-substantive sense) agents are praised, whereas in others they are blamed, Aristotle says that:

(a) in other cases praise is not bestowed, but sympathy (*suggnômê*), i.e. cases where someone does the sort of things one should not do (*ha mê dei*) because of what is such as to overstrain human nature (*tên anthrôpinên phusin huperteinei*) and such that no one could withstand. (b) But perhaps in some cases there is no such thing as 'being coerced' (*anagkasthênai*), but one should rather die suffering the most terrible things; for the things that 'coerced' (*ta anagkasanta*) Euripides' Alcmeon to commit matricide seem absurd (*geloia*). *EN* 1110a23-29

The cases of coercion mentioned in passage (a) are clearly cases of what we have called 'objectively overwhelming coercion'. But what is the connection between these cases and the ones mentioned in passage (b)? Are the cases in passage (b) a sub-class of the cases of OR-coercion referred to in (a)? It is reasonable to suppose they are (cf. *enia*, a26).

Now, the *EE* recognized cases of OO-coercion, in which sympathy (*suggnomê*) is bestowed upon the overwhelmed agent because he is subject to an impulse that is "naturally capable of over-straining nature" (*EE* 1225a22-23), where this clearly implies that the agent is regarded as having acted *involuntarily* (*akôn*, *EE* 1225a23).

Moreover, *EN* III 1 itself has admitted in its opening paragraph that "upon involuntary actions sympathy is bestowed" (1109b32). What then could be the point of Aristotle's claiming in passage (a) that in cases of OO-coercion the agent deserves sympathy (and thus acts *involuntarily*), while admitting in passage (b) that *some* cases are such that it is appropriate to say of the agent that he '*should* have died suffering the most terrible things' rather than do a morally abominable thing? For 'should have done or suffered *x*' surely implies 'could have done or suffered *x*' (i.e. 'ought' implies 'can').

One possible retort to this way of posing the question is to bite the bullet and argue that the *EN* 1110a23-29 passage suggests that cases of OO-coercion are in fact *voluntary* despite their being the object of sympathy. Nielsen for instance says that Aristotle "holds that even when the threats overstrain human nature, and lead us to do things we should not have done, the consequent action is voluntary, although we pardon the agent".[16] I think we should not be driven to such a desperate conclusion. First of all, this is a conclusion that puts Aristotle in plain contradiction with himself in the *EE* and in the *EN* itself: even though Aristotle does not think that all (bad) involuntary actions call for sympathy (e.g. *EN* 1136a5-9), he clearly thinks that all actions that call for sympathy are involuntary (*EE* 1225a21-23; *EN* 1109b32, 1111a1-2). Secondly, Nielsen's solution assumes an unjustified generalization, i.e. that *all* actions performed under OO-coercion are voluntary in spite of their being the object of sympathy. But surely Aristotle's point in passage (b) is only about *some* cases (*enia*, 1110a26).

I think it is more plausible to think that *EN* 1110a23-29 agrees with the *EE* in that the overwhelmed, coerced agent acts involuntarily, while *also* allowing for *limiting* cases – which the *EE* does not recognize. The question then is: what are the distinguishing features of these limiting cases, like the one of Alcmeon in the lost play of Euripides? What are we expected to conclude from them?

---

[16] Nielsen, K. M. (2007), p. 277. See also Klimchuk, D. (2002), p. 8; and Burnet, J. (1900), commentary *ad loc*. This also seems to be Nemesius' interpretation , who offers the example of Zeno "who bit out his tongue and spat it at the tyrant Dionysius rather than divulge secret mysteries to him", and the philosopher Anaxarchus, "who chose to be pounded to death at the hands of the tyrant Nicocreon rather than betray his friends" (Ellebaudt, N. (1565), p. 384).

We know that in Euripides' lost play, Eriphyle (Alcmeon's mother) was given the necklace of Harmonia by Polynices to persuade her husband Amphiaraus to take part in the expedition of the Seven against Thebes, which had as its aim to recover the city for Polynices. As a seer, Amphiaraus predicted his own death there, and commanded Alcmeon and his brother to avenge his future death by slaying Eriphyle their mother, invoking on them 'bareness of earth and childlessness' if they disobeyed.[17] It is plausible to think that these threats issued by Amphiaraus correspond to 'the things that coerced Alcmeon to commit matricide' (*ta anagkasanta mêtroktonêsai*, 1110a29). And notice that Aristotle says that in the limiting cases referred to "one should rather *die suffering the most terrible things* (*apothanteteon pathonti ta deinotata*)" (1110a27), thus suggesting that Alcmeon indeed killed his mother to avoid suffering a terrible death from famine and losing his progeny. How, then, can Aristotle say that the things that 'coerced' Euripides' Alcmeon to commit matricide are *absurd* (*geloia*)? Clearly, Aristotle does not mean by this that avoiding bareness of earth and childlessness is absurd or ludicrous. What he means, I take it, is that the appeal to overwhelming coercion, that is, to a fear or pain (recall the way these two are connected, as we noted in Chapter 4) that no good human specimen could be expected to withstand, is absurd, *given* the nature or significance of committing matricide.

There are two ways to understand this latter claim, depending on what the significance of committing matricide is taken to be in the example. One is the following. Recall that in Chapter 4 (Section C) I suggested that one difference between overwhelming and objective coercion was that the corresponding fear is characterized as being such that 'no one's nature could be expected to resist' *independently* of the *moral* gravity or seriousness of the instrumental action (in Alcmeon's case, matricide). We saw that this claim could be grounded (i) on the assumption that, in cases of OO-coercion, the end to be avoided is primarily represented by fear as something violently painful, and (ii) on the assumption that pain and ethical wrongness may not be *commensurable*, that is, that they cannot be weighed on the same balance, so that one can condole with S for doing *x, no matter*

---

[17] See the commentary on fr. 69 in Nauck, A. (1926). The relevant fragment in the *scholia* referring to the threat reads '*katarasmenou autou, ei mê apokteinê*[*i*], *akarpian te gês kai ateknian*'.

the moral gravity of doing x, on the grounds that the pain S had to submit to was such that no one's nature should be expected to withstand. Perhaps Aristotle is now in the *EN* denying assumption (ii), i.e. denying the incommensurability of pain and ethical wrongness, and pointing at cases (like Alcmeon's tragedy) in which the *wrongness* of the instrumental action is such that it simply *outweighs* the *pain* the agent is trying to avoid. The problem with this interpretation, it seems to me, is that it ultimately does away with objective overwhelming coercion and the distinctiveness of the plea involved in it. I can't really see how the fact that a given pain, or the corresponding fear, is humanly unbearable (which is the ground of our condolence) can ultimately depend on the degree of moral wrongness of the instrumental action.

Admittedly, one possible rejoinder is this: What if Aristotle's point here is that the human condition as such (and not just such and such contingent cultural practices and institutions) places certain constraints on the cogency of our theoretical views about what calls for sympathy? "Suppose" – so the argument would go – "that our theory says that certain situations of coercion offer a painful alternative so violent and intense indeed that any authoritative spectator would admit the corresponding fear to be irresistible and so sympathize with the agent, on the grounds that he acted involuntarily (i.e. it was not up to him to suffer the pain threatened) *regardless* of whether the instrumental action was a trifling evil or an abomination. But what if people's moral sensibility is *naturally constituted* in such a way as to make it simply impossible for them to sympathize with the overwhelmed agent when his action is morally repugnant, *albeit involuntary*?" Now, I take it that this sort of rejoinder is in line with the sort of naturalistic approach to moral responsibility found in Hume and Strawson, but I see no reason to think that this naturalistic route is open to Aristotle.

In my view, Aristotle need not deny the incommensurability of pain and wrongness. In his view, I think, Alcmeon's case is a limiting case partly because Alcmeon's plea, namely, that the fear of childlessness overstrained his human nature, collides with the fact that the horror of matricide *should have overstrained his nature as well*. Now, true as this seems to me, this cannot be the whole story about the distinctive character of Acmeon's plea, for Aristotle thinks that his plea is

'absurd' (*geloion*) and that this is a fair verdict only on the assumption that Alcmeon *should have endured* the fear of childlessness *and not committed matricide*. Now, I find it difficult to make sense of the idea that things that overstrain human nature admit of degrees (so that the horror of committing matricide *overstrains more*, as it were, than the fear of childlessness). The remaining alternative for my interpretation, it seems, is that (i) Alcmeon indeed should be *equally* overstrained in two opposite directions, in that both childlessness and matricide equally ground a legitimate appeal to OO-coercion, but this simply means that 'they cancel each other out', so to say, when it is a question of pronouncing ethically on Alcmeon's action; and (ii) to say that Alcmeon should have endured childlessness (and thus that his plea is absurd) is tantamount to saying that, *given the cancelling out* in (i), one can only judge by the relative *moral awfulness* of Alcmeon's action: it is clearly absurd (*geloion*) not to see that matricide is *the worse* of the two. Alcmeon's case is then a limiting case, but it is not a counterexample to Aristotle's theory of OO-coercion as we have developed in Chapter 4.

I conclude that there is no significant difference between the Eudemian account of objectively overwhelming coercion and the Nicomachean. Both portray OO-coercion as a plea whose role is importantly different from OR-coercion and its justificatory role. The agent in cases of OO-coercion is not *praised*, because in these cases the question is not whether he has correctly ranked the alternatives left open to him. He is rather the object of sympathy or condolence. This makes it quite clear that he *does something wrong*, albeit involuntarily. This is precisely Austin's definition of excuse.

VI

FACTUAL ERROR AND THE SOURCE OF BLAME

In the first part of this chapter (Section A and B) I provide a unified account of factual error, an account I deem necessary in view of a series of conceptual problems besetting the unity of this notion (mainly two problems: the diversity of Aristotle's examples of acting through factual error, and the wrong interpretation of what I call the primacy claim, i.e. the claim that some of the conditions of factual error are primary). In the second part (Sections C to F) I offer an interpretation of Aristotle's view on culpable factual error, which I call the 'liberal interpretation', and argue that this is the one that makes best sense of the *EN* claim that acting *di'agnoian* is *sufficient* for rendering the action involuntary (1111a16-18).

A. A general account of factual error

According to *EN* III 1, acting *di' agnoian* is one of the reasons one can offer for saying that one acted involuntarily (*akôn*) (1110b18, 1111a17, a22).[1] Now, '*di' agnoian*' has been traditionally translated as 'through ignorance' or 'by reason of ignorance' (e.g. Ross), but there are reasons to translate the phrase '*di' agnoian*' as 'through factual error' and '*agnoia*' in this context as 'error'.[2] The three terms involved in this translation, 'through', 'factual', and 'error', deserve comment. 'Factual' picks up the fact that this sort of unawareness (*agnoia*) is labelled '*he kath' hekasta agnoia*', 'unawareness of the particulars' (as opposed to '*he kathoulou agnoia*', 'ignorance of the universal', *EN* 1110b32-33). An initial reason to translate '*agnoia*' here as 'error' is that the actions done *di' agnoian* are said to be '*hamartêmata*'; 'errors' or 'mistakes' (*EN* 1135b13, 1136a7, 1110b29; *EE* 1246b3, b7).[3] A further reason is that some of the cases considered by Aristotle involve not only acting because of a lack of awareness about particular facts, but also, in Kenny's words, because of a 'positively mistaken belief'. So for example one of Aristotle's

---

[1] *EN* 1110b18-24 and 1111a19-21 say that only the one who feels subsequent pain or regret for what he has done through error acts *akôn*. I will ignore this important complexity in this chapter, because I believe it deserves separate discussion (see Chapter 7).

[2] I don't mean by this that '*agnoia*' *means* factual error, for it clearly does not mean this in '*hê kathoulou agnoia*'.

[3] See also *Poet.* 1453a9-10 (*di' hamartian tina*) and Stinton, T. C. W. (1975) on the use of this expression.

examples is that of Merope (*EN* 1111a11-12) who *believed* that her son was an enemy. What led Merope to kill the person who was in fact her son was not merely her being unaware of the fact that he was her son, but also the positive (not necessarily *false*) belief that he was her enemy.

Kenny's notion of '*positively* mistaken belief' is important, but needs further elaboration. The case of Merope can exemplify acting on a mistaken belief in different ways, for there is a distinction to be made between (1) the case in which Merope's son *is* in fact her enemy, and (2) the case in which Merope's son *is not* her enemy, in which case we would say that Merope 'mistakes her son for an enemy' (this latter is the sort of case Aristotle seems to have in mind).[4]

In 1, Merope's belief that this person was her enemy is not mistaken, in the sense that it is not a false belief but a true one. Nevertheless, merely saying that Merope *acted on* this positive belief, where the action is the one of killing her son, would be to offer a very unsatisfactory description of the situation to someone who wanted to know why she killed her own *son* (i.e. that for which she is being blamed). The excuse seems incomplete in case 1 because the fellow was in fact her enemy. There are perhaps two ways in which we can complete the description on Merope's behalf, in case 1. Her *plea* that she was unaware that this person was her son can mean: (1a) that she *believed* X to be her enemy and *not* to be her son (i.e. she acted on the belief that X was her-enemy-and-not-her-son), or (1b) that she believed X to be her enemy but *did not hold* any belief whatsoever as to whether X was also her son. 1a can be clearly characterized as a sort of 'positively' mistaken belief, but I doubt 1b can be so characterized.

2 is a clear case of acting on a 'positively' mistaken belief too: here Merope's belief that X is her enemy is simply false. In this second case, the mere fact that Merope had that false belief can be cited as an excuse for her having killed her own son. It is this sort of example that Kenny is referring to by the phrase 'positively mistaken belief'. Now, if all cases of acting *di' agnoian* are like 1a and 2, then it

---

[4] Merope's filicide was the subject of a lost play by Euripides, *Cresphontes*. Cresphontes, Merope's son, pretended to be his own murderer by disguising himself and claiming the price which had been put on his head. Merope, believing Cresphontes' story and wanting to avenge his son's supposed death, killed him while he slept. See Nauck, A. (1926), pp. 497-501.

would turn out that all cases of acting *di' agnoian* are in fact cases of acting on a *positively* mistaken belief.

But this is not the case. At least some cases of the 1-sort are better accounted for in terms of 1b, i.e. the absence of any beliefs as to whether X is F (e.g. her son), rather than in terms of 1a. Consider the example of Aeschylus at *EN* 1111a9-10. If we are to believe Henry Jackson,[5] "having been taxed with the betrayal of the mysteries" Aeschylus replied in plain prose: 'I said the first thing which occurred to me, not knowing that there was anything in it which had to do with the mysteries'. Aeschylus of course did not totally ignore that the mysteries were secret and should not be disclosed. The error was just a failure to pay attention to what he was saying. (Let us assume for the moment that cases of *culpable* factual error are also cases of acting *di' agnoian*, as I will argue). The Aristotelian description of Aeschylus' mistake so described would be that it is a case in which S acts in the *absence* of any belief as to whether the action was or was not that for which S is being blamed. When S pleads, 'I'm sorry, I just didn't think about what I was saying, I know that I shouldn't have said it', the verb 'to know' here is of course used dispositionally: it is not as if Aeschylus only *now* knows that he should not have said what he did, for he has known this all along. So Aeschylus *has* the knowledge that what he said was not permissible, a secret, etc. and he simply fails to exercise it (*EE* 1225b9-16). It is this failure to exercise this knowledge, rather than his occurrent belief that what he said was not a secret, that explains his error.

If the preceding considerations are granted, then it seems that the notion of acting *di' agnoian* encompasses both cases of acting on a *positively* mistaken belief (2 and 1a *supra*) and merely acting on a mistaken belief (1b *supra*): Aeschylus of course, *does something* on an occurrent belief, e.g. the belief that he was saying something beautiful; but this belief is not 'positively mistaken' (as in 2, or 1a, involving *false* beliefs). It is still 'mistaken', however, in the sense that it is 'partial' or 'incomplete'.

At any rate, whatever the difference between these different cases is taken to be, in both cases of positively mistaken and merely partial belief, there is an error

---

[5] Jackson, H. (1901). Jackson cites *Republic* 563 C.

or mistake that can be characterized as *epistemic*. This much, I think, is uncontroversial. Now, a more serious challenge to a unified account of acting through factual error is that the notion of acting *di' agnoian* encompasses *more* than epistemic error in the sense of acting on a positively mistaken belief, or a merely partial or incomplete one. To see this, let us consider the more controversial examples of acting *di' agnoian* offered in *EN* III 1: the catapult case (*EN* 1111a10-11) and the wrestler's case (1111a14-15). The catapult case is the case of a fellow who wanted to show how the catapult worked, and let it go off. Aristotle offers this as a case of error about *that on which* he acted (*peri ti ê en tini*), so presumably he thought (as Taylor notes)[6] that the case was one in which the fellow mistook the firing mechanism – that *on which* he put his finger – for the safety switch. So interpreted the case is one of epistemic, factual error. Interestingly enough, the catapult case can also be read in a different way. Suppose the fellow, while trying to show how the catapult worked, *inadvertently* let it go off: he didn't pay enough attention to what he was doing (where these epithets obviously imply that he is at fault to *some* extent, as we shall see). Now, this is significantly different from the rest of the examples of acting *di'agnoian*, for (so interpreted) it is not obviously a case of epistemic factual error; apparently there is no relevant *fact* that the fellow of the catapult was unaware of. The other case is that of the wrestler who, wanting only to grasp his opponent's hands,[7] struck a blow on him. This latter is a case of error about *how* (*pôs*) he touched him. Again, this latter example is significantly different from the rest of the examples of acting *di'agnoian*, for the error here is not really an epistemic one, but rather that he didn't 'weigh up his strength', or was *careless.*[8]

Another way to see that these cases and the conditions they illustrate are different from the rest, is to note that if the agent had noticed, e.g. that *what* he was saying was revealing the mysteries or that *the way in which* he was touching someone was with too much strength, the gaining of this knowledge could not have prevented the very execution of these actions (at most it would have made possible, in some cases, their interruption), since the very execution of these actions is a pre-

---

[6] See Taylor, C. C. W. (2006), *ad loc*.
[7] '*Apocheirizomenoi*' refers specifically to people who practice a special kind of wrestling in which wrestlers grasp each other's hands and are not supposed to clasp their bodies, least to hit each other.
[8] Kenny also notes this difference, see Kenny, A. (1979), p. 51.

condition of his having noticed what he was doing and how (e.g. if the wrestler had noticed that he was hitting someone with too much strength it is because he was *already* hitting someone with too much strength). This is because these two conditions (i.e. the what and the way in which) seem to be inherent to the very execution of the action, whatever this is.[9]

Now, these examples involving inadvertence and carelessness seem to show that the notion of acting *di' agnoian* is even broader than epistemic error or mistaken belief,[10] embracing also cases in which some psychological deficiency, or perhaps a character flaw, rather than a mere epistemic one, is at stake (a sort of vice or lack of skill like clumsiness, inadvertence, slips of the tongue, heedlessness, carelessness, etc.). These states or dispositions threaten the unity of the notion of factual error, because one may think that they are *directly responsible* for the resulting blameworthy state of affairs, through no sort of mistaken belief whatsoever.

I have deliberately included in the list a variety of dissimilar dispositions, and it is important to draw some distinctions. On the one hand, we have what the 19th century legal philosopher John Austin called 'states of mind supposing unconsciousness' (e.g. negligence and heedlessness). These are cases of inadvertent taking of unreasonable risks.[11] According to John Austin's definitions, one does not perform an act that one should, because one is unaware of some fact (negligence), or one does an act that one should not do because one does not advert to its probable consequences (heedlessness). Carelessness, unmindfulness, inadvertence, etc. also belong to this family. I will call these 'vices of impure agency',[12] because –

---

[9] The most complete enumeration of conditions is the one in *EN* III 1: (I) *who* (*tis*); (II) *what* (*ti*); (III) *with regard to what* (*peri ti*) or *on what* (*en tini*); (IV) *with what*; (V) *for the sake of what* (*heneka tinos*); and (VI) *how* (*pôs*). Conditions (III), (IV) and (V) are doubtless the standard conditions for Aristotle, for they are the only ones that the *EN* V 8 1135a24-26, the *EE* II 9 1225b2-6 and the *MM* 1195a17-19 enumerate. I will show below that Aristotle is inclined to regard conditions (II), (V) and (VI) as identical.

[10] Perhaps because of this, Means, T. (1927) translates '*di' agnoian*' as 'un-intelligently'.

[11] See Austin, J. (1873), pp. 240-2. Austin also includes recklessness, but I am unsure about the status of recklessness. Perhaps some cases of recklessness (e.g. in the sense defined by A. Duff as 'practical indifference', which can be displayed in someone's *failure* to notice a risk, see Duff, A. (1990), Ch. 7) are vices of impure agency, but not those in which the agent *averts* to the possible consequences (see footnote 17).

[12] 'Vices of impure agency' are the opposites of what Walker, M. U. (1993) calls the 'virtues of impure agency', defined by her (p. 241) as "corrective of temptations and deficiencies of sorts to which human beings are commonly susceptible with typically or overall undesirable results", and to be "constituted

as I will suggest – they are better thought of as dispositions, and as dispositions of an ethical nature (Austin's label, 'states of mind supposing unconsciousness' does not capture this). An altogether different family comprises slips of the tongue, clumsiness and lack of dexterity. These are cases in which one's movements or words are simply not careful enough in the sense of not being finely controlled so as to achieve one's purpose successfully, and where the resulting mischief is not involuntary in virtue of one's failure to be aware of some relevant fact about one's practical situation. In other words, the failure in these cases seems to be purely a matter of *failing to control one's body* in accordance with one's intentions.

Aristotle's view with regard to the operation of vices of impure agency like negligence (taken in this dispositional sense), heedlessness, etc. is simply that of an *epistemic error* (mistaken belief) that can be *attributed to a further fault in the agent*. Aristotle seems to think that in cases of acting *di' agnoian* that involve vices of impure agency like lack of attention, negligence or heedlessness, these faults are *indirectly* responsible for a mischief of which the *direct* causal factor is still some sort of factual *epistemic* error (*he kath' hekasta agnoia*). I will argue for this in Section D and E, so I ask the reader to take this for granted in the interim.

But what about the cases in the clumsiness family distinguished above? There are clearly cases (like Jackson's Aeschylus *supra*), in which people who utter words that 'escaped' from their mouths *know*, in a dispositional sense, that saying this would be wrong; and it is even plausible to say that, if they had paid enough attention to what they were saying and had been aware of the secrecy as a result, they wouldn't have said it. If they didn't already know in *some* sense that what they said was a secret, then their pleading that it 'escaped' from their mouths wouldn't be the relevant excuse, but rather their pleading that they just didn't know that what they said was something they shouldn't have said. Still – and this is important – in *many* of these relevantly similar cases of slips of the tongue (e.g. a native English speaker says 'I like your horse' instead of 'I like your house') it would be simply false to say that the agent said what he said *believing* something about what he was saying or, in other words, consciously saying it under *some* (true) description. If some

---

in important part by a reliable capacity to see things clearly, to take the proper moral measure of situations, so that a fitting response may be fashioned".

of Aristotle's examples are *these* sort of cases in which there is no mistaken belief to be invoked at all, then it looks as if the notion of acting *di' agnoian* has indeed no unity.

Take for instance a clear example where there is no mistaken belief to be invoked at all: through a careless movement of my arms – the sort of arm movement one does when walking – I drop a precious vase, which I am perfectly aware is there, when trying to get through a narrow corridor (without causing any damage). Aristotle would say that there is no *agency* involved in this sort of examples, if we are to base our idea of agency at least partly on the Aristotelian idea (discussed in the next section) that if an *action x* (as opposed to a mere motion) is going to be *involuntary* at all, it must not only be the case that (i) there is a true description of *x*, *TD1*, of which S is unaware (and under which *x* is involuntary), but it must also be the case that (ii) there is a true description of *x*, *TD2*, under which *x* is performed voluntarily (otherwise there is no agency involved).[13] Since (ii) is prior to (i), and no such condition as (ii) is fulfilled by the careless movement of my arms when I walk (in the vase example), then *a fortiori*, there is no factual error either.

I contend, however, that although some of Aristotle's examples may be correctly described as cases of clumsiness, none of them are such that do not involve a mistaken belief, as in the example of the vase. Cases like the one of the wrestler are not relevantly similar to the case of the vase. In the wrestler example he is only unaware of the modality of the movement of his arms, not of the movement itself (as in the precious vase example). The wrestler example is certainly discommoding because, as we have seen, Aristotle's account of it as a case in which the agent is unaware that he is moving his arms violently, does not allow one to see clearly how it can be true to say of the wrestler that, if he had been aware of it he could have avoided doing it. Because of this, the wrestler's case is better accounted for in terms, not of his not being aware of *the way in which* (*pôs*) he was actually doing something, but rather unaware of the (possible) results that his *doing it in that way* could bring about (e.g. killing his opponent). It is easy to see how Aristotle could have thought then that the *how* condition could be subsumed under the *what*

---

[13] Compare Davidson's account in his 'Agency' in Davidson, D. (2001).

condition, because unawareness as to *what* one is doing really amounts to unawareness of the action under a description that involves an unintended result. So for instance *EN* 1135b12-17 mentions the *what* condition, giving as example the person who "thought that one was not *hitting* <someone>", but does not mention the *how*, presumably because 'touching B *violently*' amounts to '*hitting* B', where the latter is a description of the former action in terms of its results (see *EE* 1225b3-4). If we put it this way, then the case of the wrestler (if it is really a case different from the precious vase example) does no longer offer any difficulties: the wrestler's case is simply a case of positively mistaken belief, a case of believing that X is A-and-not-B (i.e. believing that the action amounts to touching B violently – TD1 – and not hitting or killing B – TD2). If so, then the case of the catapult (interpreted as in p. 159) can also be interpreted as a case positively mistaken belief (this has the advantage of explaining why Aristotle did not bother to distinguish it from the case in which the operator mistook the activation bottom for the safety switch). In the next section I will return to these considerations concerning the description of the action in terms of its results.

The upshot of these considerations is that Aristotle's account of acting *di' agnoian* is a unitary account because it is the account of a sort of error which is eminently epistemic, that is, of positively mistaken factual belief or merely partial factual belief in the senses specified. But we are still not in a position to understand the force of the remaining notion in our rendering of *di' agnoian* as 'through factual error', namely, the 'through' (*dia*). The sometimes quite elusive meaning of this preposition will not become clear until Sections C-E, where I consider Aristotle's view on culpable factual error. About those cases that do not seem to fall under the heading 'factual error' (e.g. the precious vase case), let me just suggest that they seem to belong to another class of ethical ascription defeaters which do not necessarily involve agency.

B. Action-descriptions and the primacy claim

Aristotle seems to recognize that whenever I do something *di' agnoian* and thus involuntarily, I do some other things voluntarily. He notes, for instance, that "it

could be that the person being struck was his father, but *he knew only that he was a person, or one of those present, not that he was his father*" (*EN* 1135a28-30).[14] Likewise, someone may believe truly that that with which he happened to kill someone was *a drink*, but was unaware that it was hemlock (*EE* 1225b3-6). Even though Aristotle fails to explicitly draw the conclusion that it is possible for someone *to offer a drink* voluntarily, and at the same time *to offer hemlock* involuntarily, or to *strike a man* voluntarily, and at the same time to strike one's father involuntarily, it is plausible to think that this is partly what he means in these passages (see also *MM* 1188b31-36). Here 'knowingly' and 'not-knowingly' are description-relative in a relatively uncontroversial sense - namely, relative to *the degree of generality in the descriptions of the same condition* (e.g. my father, a man, a human being, a substance, etc.).

Cases of positively mistaken belief are particularly interesting in this regard, for they provoke the question of whether Aristotle recognizes, not only that e.g. a woman can offer *a drink* to a man voluntarily and offer him *hemlock* involuntarily (this being a case of merely mistaken belief), but also that a woman can *offer a love-potion* to a man voluntarily (this being a case of *positively* mistaken belief) while in fact offering him hemlock (and thus killing him) involuntarily. It seems to me that the reason why this latter distinction is more controversial is that, given that the drink *was not* in fact a love-potion, the action-description 'offering a love-potion' is simply incorrect and therefore it seems rather absurd to say that the woman *did* that action (viz. it is incorrect to say that one *offered a love-potion* voluntarily), doubtless because *doing* X implies that 'X' corresponds to a true description. Even if we prefer to say that the woman offered *a drink* to a man 'with the thought of' (*meta dianoias*, *MM* 1188b37) charming him, we still seem to be confining voluntariness to the action described as 'offering a drink'.

And clearly, Aristotle would be ready to agree that the aforementioned description-relativity is absurd. The woman in the example *cannot* be said to offer a man a love-potion voluntarily, because mere belief is not sufficient for voluntariness. What we need is knowledge,[15] and knowledge requires, at the very least, true

---

[14] Rowe's translation.
[15] *To epistasthai, to eideinai*: *EE* 1225b2, b11; *EN* 1111a23, 1135b20.

belief.[16] Notice that this does not imply that the woman gave the man a love-philtre *involuntarily* either. In fact, '*agnoia*' (*un*awareness) is eminently a privative concept, and Aristotle conceives of *agnoia* as *lack* of knowledge (e.g. *ouk eidôs EN* 1110b27, also b21). The reason why the woman offered hemlock to the man involuntarily is that (i) she offered him a drink voluntarily, and (ii) she *did not know* that the drink was hemlock, but no such reason as (ii) can be offered for her having offered a love-philtre, for the drink was not a love-philtre. Hence, the right verdict is simply to deny that there was such a thing as offering a love-philtre in the first place. As a consequence, voluntariness is description-relative only in the sense that the action-description under which what the agent did is voluntary corresponds to a *true* description of the action (however partial this could be) which is known by the agent, and involuntariness is description-relative also in the sense that the action-description under which the action is involuntary corresponds to a true description of the action, which is not known by the agent.[17]

So far so good: 'Knowingly' and 'unknowingly' seem to be description-relative in a relatively modest sense, namely, relative to the degree of generality in the description of the same condition. But what if 'knowingly' and 'unknowingly' are also relative to the true descriptions of *different* circumstances? So suppose that not only

---

[16] I think Aristotle's concept of knowledge in this context is most charitably interpreted as true belief (*doxa alêthês*). Notice that in some contexts Aristotle deems *doxa alêthês* equivalent to *epistêmê* for the purposes of his ongoing investigation (e.g. *EN* 1146b23). Admittedly, even if charitably interpreted, the knowledge condition is troublesome, because one needs to stretch this concept of true belief to its limits, not only to include *probable consequences*, but also *possible* ones. Consider a recklessness version of the catapult example, in which A lets the catapult off voluntarily in order to hit the castle and *does* advert to the possibility that the rock could hit B's house in the vicinity (although A may assume that the mischief will not ensue in the present circumstances). If one thinks that A is open to blame *for hitting B's house* (whether or not A is later shown to be *justified* in doing so), then Aristotle would have to stretch the concept of true belief, and argue that A hit the house *voluntarily*. In this section I interpret the knowledge-requirement as the requirement that the action be known to the agent under a true description, and there is perhaps a *minimal sense* of 'true description' in which A has a true belief about a consequence of his action (or the action described in terms of that consequence) if A believes there is a *possibility* that a certain effect will come about, and the effect comes about.

[17] Merely recognizing that 'knowingly' and 'not knowingly' are description-relative in the sense specified, is not satisfactory unless one has a way to block the problem generated by the free substitution of extensionally equivalent descriptions: e.g. one might still argue that, since I offered that drink to my father knowingly and thus voluntarily, and that drink is that hemlock, then I offered that hemlock to my father voluntarily. Aristotle seems to have a way to block this fallacious inference. While we would say that 'knowingly' introduces an opaque-context in which extensionally equivalent action-descriptions cannot be freely substituted *salva veritate*, Aristotle says that knowledge must not be *per accidens* (*mê kata sumbebêkos*, *EN* 1135a27, *EE* 1225b6): that is, it would not do to say that I know *x* is hemlock because I know *x* is a drink and *x* is, as a matter of fact, hemlock. In other words, I must know *x* qua (*hê*[*i*]) hemlock if it is to be said that I have offered hemlock knowingly and voluntarily. On this see Heinaman, R. (1986), pp. 131-2.

is it possible for A to offer a *drink* to B voluntarily, and to offer him *hemlock* involuntarily, but also possible for A to throw a shaft (i.e. instrument) to B voluntarily and kill B involuntarily (i.e. result). Aristotle recognizes this too, I think:

(A) it seems that the person who acts being unaware of *any* (*ti*) of the whole range of circumstances in which the action occurs and about which there can be factual error, acts involuntarily (*akôn*), (B) and most precisely, <he acts involuntarily> when the error involved the most important conditions (*kai malista en tois kuriôtatois*); and it seems that the most important of the conditions in which the action occurs are the *quid* and the *cuius gratia* (*ho kai hou heneka*).[18] *EN* 1111a15-19

Passage (A) could be taken as asserting that 'knowingly' and 'unknowingly' are also relative to the true descriptions of *different* circumstances. It is worthwhile noticing that Aristotle's talk about '*acting* involuntarily (*akousiôs*)' or '*what* the agent *did* involuntarily (*akôn*)´ (1110b19-22) without any specification as to the relevant description of 'the action' in terms of the kind of circumstance, is perfectly compatible with the fact that what the agent 'did' is involuntary relative to such description: such unqualified talk can be fully explained by the fact that blame – and let us assume, also praise – *already* picks up a particular description of the action which then becomes salient, for it is only under a particular description that the action is ascribed to an agent. Thus, one can talk simply about 'the action', as distinct from 'the action considered under this or that circumstance', in the context of blame, because this is a context in which a particular description of the action has *already* been singled out as *prima facie* blameworthy, and this description can accordingly be safely assumed by all participants.

Now, passage (B) in *EN* 1111a15-19 (*supra*) has been interpreted by Ackrill[19] in such a way that it is incompatible with the acknowledgement of the relativity of 'knowingly' and 'unknowingly' to the true descriptions of different circumstances. I must confess that I do not fully understand Ackrill's argument in support of this conclusion, but perhaps the whole issue depends on the translation of the superlative '*malista*' at 1111a17 (which Ackrill seems to take as implying that the

---

[18] I depart here from Waltzer & Mingay edition by inserting '*ho*' before '*kai*' at a18. On this see footnote 21. I give the Latin translation of '*ho kai hou heneka*' because I need to justify a certain English translation (in particular of '*hou heneka*').
[19] Ackrill, J. L. (1980), p. 98.

agent acts 'most involuntarily' when the error concerns the main conditions, and conversely 'most voluntarily' when he knows these main conditions). When Aristotle says that (A) "it seems that the person who acts being unaware of *any* (*ti*) of the whole range of circumstances in which the action occurs, and about which there can be error, acts involuntarily, (B) *kai malista en tois kuriôtatois*"; it seems that the best way to translate (B), in order to preserve the truth in (A), is "and *most precisely*, <he acts involuntarily> when the error involves the most important conditions". What Aristotle means by '*kai malista en tois kuriôtatois*', I contend, is based on the idea that amongst conditions *a, b, c, ...* f, there are some conditions, *a* and *b*, that are more important because the rest of the conditions of factual error *presuppose* factual error about *a* and *b*. Saying that factual error about *c*, about *d*, etc. presupposes error about *a* and *b* is perfectly compatible with the claim that I can act involuntarily relative to one condition and voluntarily relative to another. Passage (B) so interpreted only precludes Aristotle from saying that S can act involuntarily relative to conditions *c, d, e,* or *f* but voluntarily relative to condition *a* or *b* (the chief conditions) or conversely involuntarily relative to conditions *c, d, e,* or *f* but voluntarily relative to condition *a* or *b*.

Now, passage (B) is about the primacy of two conditions: the *quid* (*ho*) of the action, and the *cuius gratia* (*hou heneka*) of the action (1111a18-19).[20] I have suggested that these two conditions are primary (*kuriotata*), because the rest of the conditions of error presuppose them. This is what I call the 'primacy claim'. That this

---

[20] The interpretation of the text has been disputed since ancient times. All the MSS read '*en hois hê praxis kai hou heneka*' at a18-19, but this generates the following conflicts. Aristotle's preferred phrase to designate all the different conditions or circumstances about which one can be in error, is '*en hois hê praxis*' ('the things in which the action occurs', *EN* 1111a16, a24). But if this phrase is used to designate *all* the conditions specified by Aristotle thus far, then it seems rather unintelligible to say that *the most important* conditions (*ta kuriotata*) are precisely *all* the conditions in which the action is located: we should rather expect Aristotle to single out some of these conditions. And secondly, if '*en hois hê praxis*' has been unambiguously used to designate all the conditions identified by Aristotle including *hou heneka* (1111a5; 1135a26) then it is difficult to explain why Aristotle refers to this latter condition as if it was not already covered by '*en hois hê praxis*'. It is doubtless because of this that some translators (e.g. Gauthier & Joliff and Rackham) have adopted H. Richard's emendation <*ho*> at a18 and inserted a comma before it ('*en hois hê praxis, <ho> kai hou heneka*'). Besides, inserting the <*ho*> seems to be the most economical emendation. This was indeed Nemesius' (*circa* A.D. 390) interpretation in his *Peri phuseōs anthrōpou* (see Ellebaut, N. (1565), p. 129). I think Richard's emendation and Nemesius' interpretation are correct. It is worth noticing, however, that this emendation was first challenged by Aspasius, who took *en hois hê praxis* in its *specific sense* (cf. *peri ti ê en tini*, 1111a5), thus considering (along with most commentators) any emendation unnecessary. I find this latter option unviable because just two lines before at a16 '*en hois hê praxis*' is clearly meant in its general sense (and deleting it at a16 as Ramsaurer did is rather unfeasible, again for reasons of economy).

is a possible interpretation of the primacy of *hou heneka* is clear. Aspasius in his commentary *ad loc.* argues that all the conditions presuppose unawareness of *hou heneka* in the sense of 'result', 'effect' or 'outcome', as opposed to 'intention' or 'purpose', etc. Given that the second, and not the former, is the traditional meaning of '*hou heneka*', Aristotle's terminology is confusing. It looks as if he thinks that whenever S *F*s in order to Z, but instead of Z his *F*ing results in W, then one can describe the situation as 'S was unaware that his *F*ing would be/was *for the sake of* W' (see *EN* V 8 1135b12-17). But error about *hou heneka* is error about *what in fact resulted* from the action.

Consider for instance *with what* (*tini*) the agent does something. A throws what he believes to be a shaft without a sharp tip, just for the sake of exercising, believing that if he hits B with it (suppose A even believes this is likely to happen) that won't have any serious consequences. As it turns out, the shaft is fitted with a spear-point, it hits B and kills him. Why not just say that what S was unaware of was that his throwing of the shaft would kill someone or was likely to kill someone? Of course, A was unaware of this because A was unaware that the shaft had a sharp tip, but to invoke this unawareness of his is only relevant on the assumption that A didn't want to kill B. Or again, recall the example of the wrestler, which was an example of unawareness as to the *way in which* (*pôs*) he touched his opponent (whom he killed or injured). Why not say that he was unaware that his touching his opponent violently would kill him? The wrestler's plea makes sense only on the assumption that he did not want to kill his opponent. And the same applies to the other cases. The *hou heneka* (in the sense of *actual result*) is important because, whatever is the circumstance of his action that the agent is unaware of, this lack of awareness is going to be relevant only in so far as the result of what he does through factual error is something he is not seeking to bring about.[21] It is because we assume this that the agent can plead unawareness of this or that particular aspect without referring to the undesirability of the actual result.

Suppose then that a full account of cases of acting *di'agnoian* is the following:

---

[21] Which is not to say that the result is something that the agent regrets, nor that it is such that he would have brought about anyway if his factual belief had been true and complete (on this see next chapter).

168

'S was unaware that his *F*ing ____²² (e.g. 'with that…', 'on that…', 'in the way in…', etc.) which S mistakenly believed to be X, would result in Z'

(In other words, factual unawareness is only salient as a *mistake* when it results in something bad that the agent was not seeking to bring about). Notice that '*F*ing' has to stand for a 'neutral' description of an action, that is, a description that is independent, in the particular case, of the conditions about which one can be in error: e.g. 'S was unaware that *pressing with his finger* on what he believed to be the safety switch…' or 'S was unaware that his *throwing* (with) what he believed to be a mere shaft…' But then these *F*ings (e.g. pressing with his finger and throwing) are not themselves something the agent can be unaware of.

We have seen that the most viable way to regard Aristotle's primacy claim at 1111a17-19 is to take it as encompassing two primary conditions, the *quid* (*ho*) and the actual, unintended result. I believe that now I can give an account of why Aristotle believes that the *quid* of the action is primary. First of all, I believe the *kai* at a18 (*ho kai hou heneka*) is epexegetic, so that the reasons why the actual, unintended result of the action is primary are also the reasons why the *quid* is primary. Indeed, this is suggested by the fact that, of the six conditions mentioned by Aristotle, only three - (i) *with regard to what* (*peri ti*) or *on what* (*en tini*); (ii) *with what*; and (iii) *the actual result* (*heneka tinos*) - can be considered as the standard ones, for they are the only ones that *EN* V 8 1135a24-26, *EE* II 9 1225b2-6 and *MM* 1195a17-19 mention. The reason why the *quid* is not included in these lists of conditions seems to be that the *quid* of the action can be subsumed under the actual result, or vice versa. So the *EE*, after mentioning the three conditions above and repeating the first two (i.e. (i) *on what* (*hon*) and (ii) *with what* (*ô*[*i*])), replaces condition (iii), the *heneka tinos* or actual result, for (iii*) the *quid* (1225b6-7). Hence,

---

²² It is interesting to note that, with the only exception of the *peri ti* or *en tini*, the circumstances Aristotle is concerned with are captured by the type of adverbial clause that grammarians call 'adjuncts', particularly most of the so called 'process adjuncts' (of manner, of means, of instrument, of agent), adjuncts of purpose, of subject, and intensifiers (see Greenbaum, S., Leech, G. and Svartvik, J. (1972), Chapter 8). A glance at the various categories of adjuncts suggests that many of them can be added to Aristotle's list so as to arrive at a more exhaustive list of circumstances than the one Aristotle offers.

there is some textual evidence to think that Aristotle was inclined to consider these two conditions as identical. And the reason is simple: *what* (*ho*, *ti*) the action *is*, in so far as it is something the agent can be unaware of, is captured by that (true) description of it that includes its *actual result* – as we have seen, the agent cannot be unaware of it under a description that is not true – and this is precisely the sense of '*hou heneka*' that we have identified as the relevant one for our discussion.

C. Culpable Error

It is undeniable that Aristotle recognizes culpable factual error. Indeed, he develops a groundbreaking account of culpable factual error in *EE* II 9, *EN* III 5, and *EN* V 8. A necessary condition for culpable error is that the agent S does, or suffers, or omits something, that puts S in a position of mistaken belief that results in an undesirable act ('the resulting act'). The drunkard at *MM* 1195a31-34 and *EN* 1113b30-34 is an example: he drinks, as a result has hallucinations, and beats his father. But sometimes the agent does not *do anything* that puts him in an unfavourable epistemic position: this is the case of the person who is simply careless. Still, Aristotle wants to say that it is because (*dia*) of his carelessness (*ameleia*, certainly a kind of disposition, a vice of impure agency)[23] that he puts himself in an unfavourable cognitive position; he 'suffers carelessness' we may say.

A sufficient condition is provided by the fact that the agent in this sort of situation does something 'culpably' (at the very least, he is culpable for *being in*, or *getting into*, the unfavourable epistemic position). Aristotle's Ethics provides us with a more or less satisfactory definition of this adverb. (i) First, Aristotle makes it clear that what the agent is unaware of (both dispositionally and operationally) must be a 'necessary' piece of information (*anagkaion*, *EE* 1225b15), by which he probably means that the information is such that people are expected to know (both dispositionally and operationally); these seem to be the things that *EN* III 5 characterizes as 'the things that one should know' (*ha dei epistasthai*, *EN* 1113b34). If these are the things that one is expected to know, depending on one's position

---

[23] That Aristotle takes *ameleia* to be a kind of proneness or disposition not to take care, rather than as the merely legal or formal concept of negligence, is clear from *EN* 1114a2-4.

(e.g. as part of one's job, one's position as a father, or as a citizen, etc.) this requirement is perhaps too stringent in some situations, for sometimes one may be blamed for not knowing things that one does not have, strictly speaking, an *obligation* to know (given one's position or role) but for failing to know things that one could *reasonably* be expected to know or anticipate (*mê paralogôs*, *EN* 1135b17) – 'reasonably' being defined from the point of view of the virtuous spectator. (ii) Another set of considerations that Aristotle uses to define 'culpably' revolves around the question of whether it was 'easy' (*EE* 1225b15) or 'not difficult' (*EN* 1113b35) to acquire or exercise the given piece of knowledge. Even if one is expected to advert to a certain piece of information, or to have acquired it in the first place, it may have been unreasonably difficult for one to do so, or even impossible. (iii) This is not all, for Aristotle also seems to assume that *besides* this 'formal' definition of culpability, as defined by (i) and (ii), there are certain causal factors corresponding to faults in the agent that account for (*dia*) his unfavourable epistemic position: these are the vices of impure agency. The culpable agent will be someone who, at the very least, will be blamed as 'careless', 'negligent' etc. because his culpable getting into the unfavourable epistemic position expresses a faulty disposition.

Let this suffice as an account of the sufficient conditions of culpability. What I want to discuss in this and the remaining sections is the question of whether the culpable agent is or should be *blamed for the resulting act*, that is, the act that is done as a consequence of one's factual epistemic error. Aristotle's most explicit answer to this question in *EN* III 1 has seemed utterly implausible to some. *EN* III 1 maintains that acting *di'agnoian* is *sufficient* for rendering the resulting action involuntary: "As factual error is possible with regard to any of these conditions, *a man who acts through factual error with regard to any of them is considered as acting* involuntarily" (1111a16-18). Because of this claim Aristotle has been accused of being "too lenient about errors of fact" (Kenny).[24] I take it that Kenny and others think that Aristotle is much too lenient, not because they think that the sufficiency claim implies that the culpable agent is *completely* off the hook (Aristotle never says such thing), but rather because, even if Aristotle recognized that the culpable agent

---

[24] Kenny, A. (1979), p. 52. See also Williams, B. (1990), p. 4.

is *to blame for the unawareness*, he ought to make this agent *distinctly culpable also for the resulting action* (so that the agent is 'doubly hooked', as it were). On this view, Aristotle *fails* to make the culpable agent distinctly culpable for the resulting action, because on Aristotle's view:

(I) praise and blame track voluntariness (*EE* 1223a9-15, 1228a10-12; *EN* 1109b30-34; *MM* 1187a20-21), and

(II) *EN* 1111a16-18, taken at face value, says that all cases of factual error render the resulting action involuntary.

Another alternative open to those who believe that the culpably ignorant agent needs to be made distinctively culpable also for the resulting act, is to argue that Aristotle does in fact make the culpably ignorant agent thus distinctly culpable, because, although Aristotle accepts (I), claim (II) - *EN* 1111a16-18 - *is implicitly restricted to non-culpable cases.* On this view (Heinaman), Aristotle *does* think that "if A's failure to realize that X was poison were due to his own negligence, then … *A voluntarily poisoned B*."[25]

Notice that both the former view (Kenny) and the latter (Heinaman) take seriously Aristotle's claim (I) that praise and blame track voluntariness. But someone may argue that Aristotle is inclined to *deny* (I) and accept (II), and still hold that Aristotle *does* make the culpably ignorant agent distinctly culpable also for the resulting act: that is to say, in cases of culpable factual error the agent does the resulting act involuntarily, but he can still be blamed *for it*, because praise and blame do not track voluntariness. This view is less plausible (in fact, no one has offered this interpretation, to my knowledge), but it is still worth taking into account for the sake of exhaustiveness. So, let us say that Kenny's and Heinaman's view, and this latter possible view all make the following assumption ('CA' for 'conservative assumption'):

---

[25] Heinaman, R. (1986), p. 134. My italics. See also Roberts, J. (1989), p. 24. Kenny also considers this alternative, but he is hesitant (*ibid*. p. 52-53).

CA: In those cases where the error is not excusable, the act done as a result is not excusable either

If the resulting act (e.g. beating one's father or letting the catapult off) is not excusable, and Aristotle *does* in fact endorse CA, then either (a) this is because he thinks that the resulting act is not involuntary, but rather voluntary (that is, Aristotle maintains (I), but denies (II)); or (b) because the resulting act is involuntary, but still blameworthy (that is, Aristotle maintains (II), but denies (I)). Let me call (a) and (b) the 'conservative interpretation' (of Aristotle's view on culpable factual error). Let me call (a) the 'voluntary version' of the conservative interpretation, and (b) the 'involuntary version'.

In the remaining sections, I want to argue that the conservative interpretation, in any of these two versions, is wrong, because both the claim that (I) praise and blame track voluntariness and the claim that (II) all actions done through factual error are involuntary, should be taken at face value, and further, that Aristotle never abandons these claims. Finally, I also want to argue (against Kenny and others) that Aristotle *need not* hold CA in order to recognize cases of culpable error (and accordingly, he *ought not* to abandon any of these two claims, (I) or (II)).

On the contrary, some of Aristotle's remarks suggest that he maintains what I call 'the liberal view'.[26] According to this view, CA is false, and the culpable agent is *only blamed* for his getting into/being in the epistemically defective state, and not for the resulting act, which is involuntary and therefore excusable (given (I) and (II)). Finally, I will be suggesting the liberal view about culpable error is not necessarily too lenient.

D. *EE* II 9

As we have seen, *EE* II 9 offers an interesting account of culpable error:

---

[26] In using the labels 'conservative' and 'liberal' I follow Smith, H. (1983). I have been greatly benefited from this article in trying to get a grasp of the deep philosophical motivations underlying these two views.

All the things <that someone does> in error (*agnoôn*) and through error (*dia to agnoein*), <he does> involuntarily (*akôn*). And since knowing and understanding are of two sorts, one being the possession (*to echein*) and the other the use (*to chrêsthai*) of knowledge, there is a way in which it may be fair (*dikaiôs*) to say of the person who has the knowledge but is not using it that <he acts> in error (*agnoôn*), and there is a way in which it may be unfair (*ou dikaiôs*) <to say this>, e.g. if he fails to use his knowledge because of carelessness (*di'ameleian*). And similarly someone would be also blamed (*psegoito an*) for not having (*mê echôn*) the knowledge, if it were something that was easy or necessary and his not having it was due to carelessness or pleasure or pain (*di' ameleian ê hêdonên ê lupên*). *EE* 1225b10-16

When Aristotle says that "all the things that someone does *in* error (*agnoôn*) and *through* error (*dia to agnoein*), he does involuntarily (*akôn*)", he seems to be considering 'in error' and 'through error' as different conditions (the *kai* at b10 does not seem to be epexegetic). The question is: what is the relation between the two, according to the *EE*? Is it for instance that (a) everyone who acts in error (*agnoôn*) acts in a state of factual unawareness, but not necessarily involuntarily, unless *he also* acts through error (*dia to agnoein*), where (b) the 'through error' phrase (*dia + to agnoien/agnoian*) is used to emphasize the fact that there is no further causal factor beyond S's epistemologically defective condition for which S can be blamed? This would support the voluntary version of the conservative interpretation, for it would mean that acting in error (*agnoôn*) but *not* through error (i.e. not *dia to agnoein/agnoian*) corresponds to those cases not covered by the *EN* 1111a16-18 claim that acting through error (*di' agnoian*) is sufficient for involuntariness. I want to show that, whereas (b) seems true, (a) is open to doubt, for *EN* V 8 (= *EE* IV 8) shows that actions done in error (*agnoôn*) but not through error (not *dia to agnoein/agnoian*) are *involuntary*:

Of involuntary actions some are such as call for sympathy, others not. All the mistakes that people do (*hamartanousi*) not only in error (*agnoountes*) but also through error (*di'agnoian*), call for sympathy (*suggnômonika*); on the other hand when they act not through error (*mê di' agnoian*) but in error (*agnoountes*), because of a state of feeling (*dia pathos*) that is neither natural nor human, sympathy is not called for. *EN* 1136a5-9[27]

---

[27] Rowe's translation with slight modifications.

This passage is interesting because it makes clear that someone may be *in* (epistemic, factual) error, but act *not* through error but rather through (*dia*) something else, e.g. through an overwhelming passion. If this is the case, Aristotle here insists on the fact that, even though the resulting act would *not* call for sympathy, it would still count as something involuntary (*akousion*). Not all actions that are done *in* error (*agnoôn*), then, are 'involuntary-*through* error', so to say: some are done involuntarily-through other factors, and it is natural to think that these other factors are carelessness (*di'ameleian*), pleasure (*dia hêdonên*) or pain (*dia lupên*), as in *EE* II 9. What then does 'through x' ('*dia x*)' add to 'in error' ('*agnoôn*)'? The connection between these expressions, as employed by *EE* II 9 and *EN* V/*EE* IV 8, is that whereas merely acting *in* error is acting unaware of this or that practical circumstance, and *involuntarily*, adding 'through error' (*di' agnoian*) by way of emphasis signals the absence of any ethically relevant causal factor (i.e. a vice of impure agency) responsible for one's factual unawareness (for which one can be blamed); and adding 'through passion' (*dia pathos*), 'through carelessness' (*di' ameleian*), 'through pleasure' (*dia hêdonên*) or 'through pain' (*dia lupên*), signals the presence of a further (ethically relevant) causal factor responsible for one's factual unawareness. *EN* 1136a5-9 emphasizes the fact that, when these further causal factors are present, they are dispositions of the agent S the salience of which does not call for sympathy, but rather for blame (i.e. S is called 'careless', 'impetuous', etc.), even though the resulting act is *involuntary.* This clearly shows that the voluntary version of the conservative interpretation is wrong; in cases of culpable error S is open to blame because of the operation of a vice of impure agency, not because of the voluntariness of the resulting act, which is in fact involuntary. The only peculiarity of these passages of *EE* II and *EN* V/*EE* IV is that Aristotle is here using '*agnoôn*' to mark (a) *all* cases of involuntariness due to factual error, and '*di'agnoian*' to mark (b) their non-culpability. Conversely, in *EN* III 1 - as we shall see in Section E - Aristotle uses '*di'agnoian*' to mark (a) – and seems to have no way of designating (b).

These considerations concerning the *EE* II 9 and *EN* V 8/*EE* IV 8 use of the expressions 'in error' ('*agnoôn*)' and 'through error' ('*di'agnoian*') are very

significant, for they allow one to understand the force of Aristotle's claim at 1225b11-16 that "it would be fair (*dikaiôs*) to say of the person who has the knowledge but is not using it that <he acts> in error (*agnoôn*), but unfair (*ou dikaiôs*) <to say that he acts in error>, e.g. if he fails to use his knowledge because of carelessness (*di'ameleian*)". By saying that it would be unfair to say that the agent who fails to activate his knowledge because of carelessness 'acts in error', Aristotle simply means that the agent is blameworthy *for something*: in other words, merely saying that the careless agent acts in error is *misleading*, not because false, but rather because it suggests, without further qualification, that he is simply let off the hook because of his unawareness (i.e. he acts *agnoôn di'agnoian*), when in fact he is to blame for (*at least*) his carelessness (i.e. he acts *agnoôn di' ameleian*). All that follows from his acting 'in error through carelessness' is that what is ascribed to the culpably ignorant agent, is his carelessness. Indeed, the observation that follows, that is, that "someone would be also blamed (*psegoito an*) *for not having* (*mê echôn*) the knowledge, if it were something that was easy or necessary and his not having it was due to carelessness or pleasure or pain" (b14-15) shows that the blame is clearly directed at the agent's failure to have the knowledge (*mê echôn*), rather than the resulting act. This makes it natural to think that the failure *to use* knowledge at b13-14 is also the target of blame, rather than the resulting act.[28] This is not only

---

[28] This is important. One strategy available to the conservative, it seems, is to understand 'culpably' in terms of what one could call a 'broad notion of factual knowledge' ('b-factual knowledge'). It is evident to all parties in the controversy that the drunkard does not know that this person is his father just on the grounds that it would have been easy to know this if he were not drunk. But Aristotle does say that the person who *has* this knowledge dispositionally, and has not enacted it, knows *in a way* that this person is his father. If one understands b-knowledge as the knowledge that S possesses and is inactive, but that S could easily enact, then it is not that clear why the conservative interpreter cannot use this notion of b-knowledge to show that, at least in some cases of culpable factual error the resulting act is voluntary in virtue of b-knowledge. Why not say, for instance, that the drunkard does know (although in a dispositional way) that this person is his father, and thus that he beats his father voluntarily? If this were Aristotle's view, then it would be reflected in *EE* 1225b11-16, but this is not what Aristotle is saying here. He makes it clear that it is *fair* (or 'correct') to say that the person who fails to activate his knowledge because of carelessness 'acts in error' (where this means that he acts involuntarily, as I have shown). This shows that b-factual knowledge is not sufficient for voluntariness. Moreover, the only place where Aristotle uses a notion of dispositional knowledge to justify the voluntariness of an action is in his discussion of incontinence in *EN* VII, yet here the dispositional notion of knowledge is not b-factual knowledge, but rather (I take it) something like the sort of knowledge involved in the Epistemic Closure Principle (viz. If S knows P, and S knows that P entails Q, then S knows Q). If (a) S *knows* that all sweet food is unhealthy, and (b) S knows that this is sweet food, then (c) S knows that this is unhealthy (*EN* 1147a24-32) even if S has not actually connected the universal premise in (a) with the particular premise in (b). This is why the impetuous incontinent eats

evidence against the voluntary version of the conservative interpretation, but also evidence against its involuntary version, for the blame here is tracking voluntariness (i.e. one's voluntarily getting into the unfavourable epistemic position).

Nevertheless, *EE* II 9 cannot be so easily disposed of, for in this chapter the discussion of culpable factual error is concluded with the words '*taut' oun prosdioristeon*' (1225b16), in Rackham's translation, "these points must be added to our definition <of the voluntary>". Apparently, this would have the implication that, in order to decide whether an action is voluntary or not one must *also* consider whether the agent was indirectly responsible for it, that is to say, whether there was a further vice of the agent that was responsible for his unfavourable epistemic position. Aristotle's concluding words, however, need not have such an implication. First, notice that Aristotle only says that these considerations about culpability must be 'added' to the previous definition of voluntariness at 1225b8-10 (which does not include the culpability conditions): he does not say that this definition should be *amended* (which is what the voluntary version of the conservative interpretation needs). Furthermore, on the liberal interpretation, there is an obvious sense in which these culpability conditions have to be 'added' to the definition: for a definition of voluntariness is a definition of those conditions under which agents are open to evaluation, and the culpability conditions are part of these conditions, in so far as the culpably ignorant agent *is blamed* - as the liberal claims - for getting himself into the epistemically defective condition (e.g. for his lack of concern for the risks, his impetuosity, etc.).

Now, I have argued in this section that according to *EE* II 9 and *EN* V/*EE* IV 8, to say that S acts (or 'errs') in error (*agnoôn*) *and* through error (*di' agnoian*) is to say, by way of emphasis, that there is no (ethically relevant) causal factor contributing to S's factual unawareness for which S can be blamed, whereas to say that S acts in error and through *x* (*agnoôn* and *dia x*) - where '*x*' ranges over a series of ethically relevant factors like pleasure, pain or carelessness - is to say that there is something *x* (which is *not* S's *agnoia*) beyond S's factual unawareness for which S

---

unhealthy food *voluntarily* (*EN* 1152a16-17). This is not the sort of b-factual knowledge the conservative needs, however.

can be blamed. When S acts in error (*agnoôn*, in the sense used by *EE* II 9 and *EN* V/*EE* IV 8) S's factual error is still the direct cause of S's erring in the way he does, thus rendering the resulting act involuntary. (Indeed, it was certainly natural for any Greek reader to take the causal meaning of the participle '*agnoôn*' for granted). Further, Aristotle makes it clear in these texts that the only target of blame is his epistemically defective condition, or that factor x that explains (*dia*) such condition. Nothing in what Aristotle has said so far supports the conservative interpretation, in any of its two versions.

Now, suppose '*di' agnoian*' were used in *EN* III 1 in the same sense in which it is used here in *EE* II and *EN* V/*EE* IV, namely, as signaling the fact that there is no further factor that is responsible for the agent's epistemically defective condition. If this was the case, then the conservative could argue that the *EN* III 1 claim that acting *di' agnoian* is sufficient for acting involuntarily (1111a16-18) is implicitly restricted to non-culpable cases. The next step in my argument is to show that this is not the case, because here in *EN* III 1 Aristotle's use of '*di' agnoian*' and '*agnoôn*' is different from his use of these expressions in *EE* II 9 and *EN* V/*EE* IV 8.

E. *EN* III 1 and 5

EN III 1 (1110b24-1111a2) also draws a distinction between acting *through* factual error (*di' agnoian*) and acting *in* error (*agnoôn*), examples of this latter case being the agent who is drunk or angry. One may think these agents are described as acting *in* error because they are responsible for acting through factual error (as in *EE* II 9 and *EN* V/*EE* IV 8).[29] But this is not so. Just after drawing the distinction (a) between acting *through* factual error and acting *in* error, Aristotle goes on to distinguish (b) between *universal* unawareness (*he kathoulou agnoia*) and *factual* unawareness (*he kath' hekasta agnoia*), and nowhere in the text does Aristotle suggest that these two distinctions, (a) and (b), are distinct. On the contrary, the transition from the 'through factual error'/'in error' distinction to the 'factual unawareness'/'general unawareness' distinction is marked by the phrase '*agnoei*

---

[29] As Kenny does in Kenny, A. (1979), p. 49

men oun'[30] (1110b28), making it natural to think that the two distinctions are to be taken as equivalent. Accordingly, Aristotle's point here seems to be that the agent who acts *in error* is not exculpated but rather blamed, not because he is responsible for the situation in which he acts through factual error – though he may be so responsible anyway – but rather because he is ignorant of (viz. he acts *in error* with regard to) the universal, normative premise of the practical syllogism, e.g. someone whose drunkenness or anger makes him believe at that time that it is alright to beat his own father.

Where Aristotle *does* consider in the *EN* the case of people who are responsible for being in an epistemically defective condition which resulted in a wrong act, is at *EN* III 5. Here he notes that legislators "impose corrective treatments and penalties on anyone who does wrong actions, unless his action is forced or is *caused by an unawareness that he is not responsible for*" (1113b24-26). Aristotle's apparent endorsement of this practice may suggest that, on his view, the act that results from factual error is not excusable (either because it is voluntary, or blameworthy, its involuntariness notwithstanding) when the error is culpable. Aristotle does not think so, however. He goes on to note that legislators "impose corrective treatments *for the unawareness itself* (*ep' autô*[*i*] *tô*[*i*] *agnoein*), if the agent seems to be responsible for the unawareness. A drunk, for instance, pays a double penalty; for the principle is in him, since *he controls whether he gets drunk*, and his getting drunk causes his unawareness" (1113b30-34). Granted, arguing that one acted through a factual error for which one was responsible does not let one off the hook, but Aristotle is not saying that this is because, *when* one acted, one acted *voluntarily*, nor is he saying that the resulting act is blameworthy. The point Aristotle stresses (*ep' autô*[*i*] *tô*[*i*] *agnoein*, *EN* 1113b30) is rather that, in these cases, one is punished (or blamed) for the action that *caused* the unawareness (e.g. getting drunk) and thus *for the unawareness itself* that caused the crime, rather than for the crime itself (i.e. the action caused *by* the factual unawareness).[31] This is the same point that the *EE* makes by maintaining that in cases of culpable error, the agent is

---

[30] 'Therefore, on the one hand…'
[31] For a similar observation, see Charles, D. (1984), pp. 256-258.

blamed "for not having (*mê echôn*) the knowledge" (*EE* 1225b14), and I think it is pretty strong evidence in favour of the liberal interpretation.

The conservative interpreter may insist on Aristotle's remark that in these cases the penalties are *doubled* and argue that this should be taken to suggest that the drunkard is responsible *both* for getting drunk *and* for what he does when he is drunk. This is in fact the best evidence that the conservative interpretation (in any of its two versions) can adduce in its support. Aristotle, however, does not seem to mean this. In the *Politica*, where the law of double penalty is attributed to Pittacus, Aristotle observes that "a special law of his is that if men commit any offence when drunk they are to pay a larger fine than those who offend when sober" (1274b19-21), and his account of the law of double penalty is that double punishment was use to deter people from getting drunk on considerations of mere expediency (*to sumpheron*), and *not* because drunk people were responsible for doing wrong *twice*. The meaning of '*to sumpheron*' is clearly explained by Diogenes Laertius in his biography of Pittacus: "Among the laws which he made is one providing that for any offence committed in a state of intoxication the penalty should be doubled; his object was *to discourage drunkenness* (*hina mê methuôsi*), wine being abundant in the island."[32] This textual evidence shows that there isn't strong evidence that Aristotle thought the penalty was doubled on the ground that there were two offences.

F. Concluding remarks

I believe then that there is no evidence for thinking that Aristotle ever abandons or qualifies the *EN* 1111a16-18 claim that the person who acts through factual error (in the *EN* III sense of *di' agnoian*) with regard to any of the relevant circumstances of his action, acts involuntarily. Nor is there any evidence to think that he abandons or qualifies his view that praise and blame track voluntariness. I think it is reasonable to conclude that, on the contrary, Aristotle holds a liberal view about cases of culpable factual error. That is, he thinks that even when S is responsible for

---

[32] *Vitae* I. 76. Hicks' translation.

generating the 'unfavourable' conditions (perhaps not only the epistemic ones, but also the conditions of violence?) that result in his doing a wrong act, it is still appropriate to say that the resulting act is involuntary, and appropriate to say that it is not blameworthy.[33] Is this view too lenient or otherwise implausible? I don't think it is. As we have seen, the liberal view on culpable error has to deny the following assumption:

CA: In those cases where the error is not excusable, then the act done as a result is not excusable either

But all the liberal interpretation of Aristotle is committed to understand by 'not excusable' is 'open to (Aristotelian) blaming *logoi*'. Aristotle denies CA because he thinks that factual error is sufficient for involuntariness, and blaming *logoi* track voluntariness. Aristotle need not deny, however, that in cases of culpable factual error culminating in an actual bad result the culpable agent is distinctly 'culpable' *on non-ethical grounds* for the resulting act (nor need he deny, accordingly, that these cases can be distinguished from those that do not culminate in an actual bad result on non-ethical grounds, e.g. for the purposes of the application of the criminal law). Hart & Honoré offer an example in which the accused "brutally assaulted and raped the deceased who thereafter swallowed 6 tablets of bichloride of mercury. She died either from this poison or from the poison and the wounds inflicted by the accused".[34] As they observe, one might have expected the judges to have tried to answer the question "whether or not the defendant's [causal] force had continued to operate harmfully until death occurred, or whether or not it had come to rest in a position of apparent safety", but instead they tried to answer the question "whether it was *socially advantageous* to give legal effect to the relation between the

---

[33] If I am right and this is indeed Aristotle's view, then there are grounds for thinking that 'acting voluntarily' or 'involuntarily' are *non-historical* notions for him, at least in the sense that, whether S's act is *hekousion* or *akousion* is not determined by features of the act's causal history. For responsibility as a historical notion see Fischer & Ravizza (1998), Chapter 7. I think that responsibility as a historical notion is closely connected to responsibility as accountability. Fischer & Ravizza make this clear (see p. 187 and pp. 194-200). The Strawsonian view (which they endorse, *ibid*. pp. 5-8) for instance, is that S-reactive attitudes are checked as soon as their potential target is seen as someone who has had no control whatsoever over the very causes and origins of his own desires, i.e. as someone whose conative structure has been the mere result of its having been developed under unfortunate circumstances.

[34] Hart & Honoré (1985), p. 95.

defendant's acts and the death".[35] The case is structurally similar to the Pittacus example and to Pittacus' own justification of double-penalty in terms of expediency (*to sumpheron*).

More significantly, these are not the only grounds for 'distinctive culpability' for the actual bad result that Aristotle need not deny. Certainly the strongest support for the conservative view on culpable error (now not just as an interpretation of Aristotle), according to which the actual result substantially contributes to the overall blameworthiness of the culpable agent, is derived from cases of moral luck (in particular, from examples of what Nagel calls 'luck in consequences').[36] For instance, Nagel argues that "if the driver was guilty of even a minor degree of negligence – failing to have his brakes checked recently, for example – then if that negligence contributes to the death of the child, he will not merely *feel terrible*. He will *blame himself* for the death. And what makes this an example of moral luck is that he would have to blame himself *only slightly* for the negligence itself if no situation arose which required him to brake suddenly and violently to avoid hitting the child. Yet the negligence is the same in both cases, and the driver has no control over whether a child will run into his path".[37] Or again: "If someone has had too much to drink and his car swerves on to the sidewalk, he can count himself morally lucky if there are no pedestrians in its path. If there were, *he would be to blame for their deaths*, and would probably be *prosecuted for manslaughter*. But if he hurts no one, although his recklessness is exactly the same, *he is guilty of a far less serious legal offence and will certainly reproach himself and be reproached by others much less severely*".[38]

Nagel here clearly employs a sense of 'blame' as the expression of a sanctioning disposition, which is the sort of blame peculiar to accountability. If *all* the (modern) conservative claims is that the culpable agent *deserves more reproach* in virtue of the actual outcome, in that his moral community reacts (or believe should react) with more negative S-reactive attitudes, feel more intensely about it, etc., then, I think, the conservative intuition seems compatible with an Aristotelian,

---

[35] *Ibid*. p. 99. Italics are mine.
[36] Nagel, T. (1997), p. 29.
[37] *Ibid*. p. 29. Italics are mine.
[38] *Ibid*. p. 29. Italics are mine.

liberal notion of culpable error. The enhancement of blameworthiness comes into the picture, as it is clear in Nagel's paper, in the form of guilt, self-reproach and other people's reproach, and punishment: all of these being practices within the family of the S-reactive attitudes. Perhaps, then, the actual consequences of the culpable act *do* bring with them an increase in S-reactive – other-regarding and self-regarding – attitudes and/or the belief that it should be appropriate to react or sanction more severely when there is an actual upshot of the culpable act than when there is not. But I see no reason why Aristotle, on the liberal interpretation, is committed to deny this. Aristotle's view on culpable factual error, then, is not too lenient, nor is it rendered implausible by our intuitions about cases of moral luck.

VII

THE PAIN CONDITION

Aristotle claims that forced motion is painful, and he claims that involuntary actions done through factual error are also painful, provided the agent has learned the relevant facts. This is what I call the pain condition. In this chapter I will interpret the pain condition on the basis of Aristotle's claim that the V-agent does the V-act without the relevant sort of pain. It follows from this that someone who does the V-act and has the relevant sort of pain, is not a V-agent (or is a non-V agent). I develop what I think is the appropriate notion of pain and lack of pain as moral aversion- and compliance-attitudes and argue that central to these attitudes is their focus on the ethically salient features of an action. I argue that the pain and lack of pain relevant to the pain condition are moral aversion- and compliance-attitudes, because involuntary behaviour can also have ethically salient features.

A. The pain condition and the involuntary

As we have seen in Chapter 3, the pain condition figures in Aristotle's account of violence (*bia*): Aristotle claims that what is done through violence (*bia*[i]) is painful.[1] And as we shall see in what follows, the pain condition also figures in his account of factual error (*di' agnoian*): Aristotle *seems* to suggest that an action done through factual error is only involuntary (*akousia*) if the agent subsequently regrets it (*EN* 1110b18-24, 1111a19-21). In spite of the obvious differences in the way the pain condition is connected to factual error and to violence (differences of a temporal order), there are reasons to think that it is the same condition Aristotle is referring to in both accounts, carrying the same ethical significance. In fact, Aristotle wants to say in the *EN* that *the involuntary* (*to akousion*) is painful (*EN* 1111a32). This latter claim suggests indeed that the pain condition is applied uniformly across the different sorts of involuntary action, but it is misleading – or so I shall argue – in that it suggests that the pain condition is a condition that affects the status of the action, and thus that it is relevant for ethical ascription only in so far as *the actions* as affected by this condition are *involuntary*. In this section I want to argue that, on the contrary, the pain condition is relevant for ethical ascription, not because it affects

---

[1] *MM* 1188a2-3; *EE* 1224a30-31, 1224a36-8, 1224b15-21; *EN* 1110b12; *Phys.* 230b3-8; *Met.* 1015a26-33; *Rhet.* 1370a11-12.

the voluntary status of the action, but rather directly affects the voluntary status of *the agent*.

We saw in Chapter 3 that the pain condition and the contrariety condition, according to which behaviour brought about through violence is contrary to the organism's impulse (*hormê*), are mutually entailing. Nonetheless, we also saw that Aristotle is committed to recognize a core notion of violence, one based solely on the externality condition, so that a necessary and sufficient condition for violent motion is that the source of that motion be external to its subject. So the question arises: What is the point of the robust definition of violence, the one taking the contrariety and pain conditions into account?[2] I want to argue that the best way to understand this robust notion of violence is to see it as conjoining two different conditions into one, the pain condition, affecting the status of the agent, and the externality condition, affecting the status of the action; and that it is better to regard the pain condition as a separate condition of ethical ascription – i.e. to regard pain as directly affecting the status of the agent.

I think the same can be said about factual error, so let me start by justifying these claims with regard to it. Just as with cases of violence, there is the case of someone, e.g. Aeschylus, who sincerely regrets (a regret which Aristotle explicitly equates with pain, *EN* 1110b19) having revealed the mysteries, and there is the case of someone who does not regret having revealed them at all. In the *Ethica Nicomachea* the Stagirite distinguishes these two 'types' of acting *di'agnoian*:

What comes about through factual error is all *ouch hekousion*, whereas what is *akousion* is what causes pain and involves regret; for the person who has done (*praxas*) whatever action through factual error, if he is not displeased at his past action, has not acted (*ou pepraxen*) *hekôn*, in so far as it was something he didn't know <he was doing>, but he has not acted *akôn* either, in so far as he is not pained <at his having done it>. What comes about through factual error, then, seems to fall into two types: someone who feels regret (*ho en metameleia*[*i*]) seems to be *akôn*, while the one who does not feel regret – well, since he is different from the former, let him be a *ouch hekôn* agent';[3] for since he is different, it is better that he should have a name to himself. *EN* 1110b18-24

---

[2] See the definition V2* in Chapter 3, Section C: *X*'s motion or rest is by violence iff (1) its source is in something external to X's *whole motivational set* and (2) is against *X's* natural tendency (*hormê*).
[3] I follow here most commentators in taking the negation '*ouch*' with '*hekôn*' rather than with '*estô*' (the same in a18).

Aristotle's use of '*ouch hekousion*' (lit. 'non-voluntary') and '*akousion*' ('involuntary') in this passage is unprecedented (this is why I have left them un-translated, for the purposes of this section), and it equips him to make a crucial distinction. To see this, however, it is important to resist one interpretation of this passage. On this interpretation, Aristotle is arguing that there is "a tripartition of actions into the voluntary, the non-voluntary [*ouch hekousion*], and the involuntary [*akousion*]",[4] the latter two being his focus in this passage. On this interpretation, Aristotle is using '*ouch hekousion*' only as *one of two* categories *of actions* due to factual error, namely, the one encompassing those actions due to factual error that are *not* subsequently accompanied by pain, and using '*akousion*' simply as the other, namely, the one encompassing those actions due to factual error that *are* thus accompanied by pain.

`   But this is not what Aristotle is saying. First, he is saying that *every action* that is the result of factual error is *ouch hekousion* (b18). '*Ouch hekousion*' here designates the *genus* of actions due to factual error that are involuntary *regardless of whether* they are or not subsequently followed by pain. *This* genus is the one that contrasts with the voluntary (*hekousion*), so that, strictly speaking, there is no tripartition of actions, but a bipartition: *hekousion* and *ouch hekousion*. Now, someone may object to this 'bipartition view' that it is incorrect, because later in the passage Aristotle uses the phrases '*ou peprachen hekôn*' ('has not acted voluntarily' or 'has acted non-voluntarily') and '*ouch hekôn*' (1110a20-21 and a23 respectively) to designate the *action* due to factual error that is *not* subsequently followed by regret (i.e. the action done 'non-voluntarily').[5] If this is correct, then the bipartition view seems wrong, because it would imply that Aristotle is just using the *same phrase*, '*ouch hekousion*' to designate both the genus (i.e. what is due to factual error regardless of whether it is subsequently followed by pain) and one of its

---

[4] Sauvé-Meyer, S. (1988), p. 129. Her interpretation is partly supported by her taking the *ouch* at a18 with '*estin*' ("is not voluntary", *ibid.* p. 127) rather than with '*hekousion*' ("is non-voluntary", as Rowe translates it). The latter is much more plausible given the position of the words.

[5] Sauvé-Meyer, for instance, reads '*ouch hekôn*' at a23 adverbially ("[he acts] non-voluntarily", *ibid.* p. 128), and the same with '*akôn*' at a21.

species (i.e. what is due to factual error and not subsequently accompanied by pain); but this is surely implausible.

In spite of these appearances, this is not what Aristotle is doing with these phrases. First, Aristotle is using '*ouch hekôn*' to designate, not the action, but *the agent* (i.e. non-voluntary agent) who does not feel subsequent regret for having acted through factual error, when learning the relevant facts (*ho mê metamelomenos*, b23); likewise, he is using '*akôn*' the designate *the agent* who feels subsequent regret for having acted through factual error (*ho en metameleia*[*i*], b22).[6] Accordingly, in my view, Aristotle is: (i) using '*ouch hekousion*' to qualify *all* those *actions* that are involuntary through factual error, *regardless* of whether they are followed by pain and (ii) using '*akôn*' and '*ouch hekôn*' to characterize *the agent* on the basis of his attitude towards that involuntary action. He can do this because the Greek terms '*akôn*' and '*hekôn*' are only predicated of agents, and agents can be 'open' or 'closed' to evaluation.

Secondly, in a way Aristotle is also pointing at the voluntary status of the *agent* when he says that "the person who has done (*praxas*) *x* through factual error, if he is not displeased at his having done *x*, has not acted (*ou pepraxen*) *hekôn* … but has not acted *akôn* either, in so far as he is not pained at his having done x" (b19-22). Admittedly, Aristotle is here using '*akôn*' and '*ouch hekôn*' to qualify actions (i.e. as adverbs). But notice his careful use of the perfect.[7] Aristotle is not saying that the person having or lacking regret about a past action done through factual error, *did* something *akôn* or *ouch hekôn.* This is presumably because the involuntary status of *what the agent did* (aorist) is immune to the subsequent negative or positive attitudes towards it. We could say that in cases of acting through factual error what the agent *did* is involuntary (genus) regardless of how he happens to feel about it. By contrast, I think that the use of the perfect tense to refer to the action now indicates, as Aristotle makes clear, that the *present* status of the *agent* has been affected by his subsequent attitude towards his *past* action (aorist).

Let me insist on this. It is tempting for an English speaker to think that the perfect in Aristotle's Greek corresponds to a *past tense*, but this would be a hasty

---

[6] Ross' translation captures this nicely.
[7] On Aristotle's use of the perfect in this connection see also Broadie, S. (1991), pp. 139-140.

assumption to make.[8] For whereas it is true that 'S has *F*ed' can be merely used to report past experiences (as in 'S has been to Chile a couple of times'), thus referring to past occasions of *F*ing, it does not have to be so used. This is because of the importance of the aspect in Greek. The perfective aspect, as its name in Greek and English implies, indicates "that the action or state to which it refers is complete; 'I have built the house' implies that there is no more of the house left to build, while 'Johnny has been a good boy' implies that Johnny has left unfulfilled none of the conditions necessary for being a good boy".[9] Thus understood, sentences of the form 'S has *F*ed' are sentences that can be true only in virtue of present conditions. In the case of 'S has *F*ed *akôn/ouch hekôn*' these present conditions include the agent's present attitude towards his past action. The use of the perfect then indicates the fact that we *now* judge the agent differently with regard to his past action, because of his present attitude towards it: 'what the agent has done' (as opposed to what he did) is then a complex whole encompassing what he did (the isolated piece of behaviour that is now gone) and our subsequent judgement (our judgement now) of the agent as '*akôn*' or '*ouch hekôn*' on the basis of his present attitudes towards it.

Accordingly, I don´t see any serious objections to the bipartition view. On this view, Aristotle is saying that (i) there is a generic category of *actions* (*to ouch hekousion*) that are *involuntary* (I see no problem in so translating '*ouch hekousion*') due to factual error regardless of the agent's subsequent attitude towards them, after he has learned the relevant facts; and (ii) there is a distinction between *agents* ('*ouch hekôn*' and '*akôn*') depending on their subsequent attitude towards them. This having been established, it is also important to my argument that these two claims, (i) and (ii), hold true of involuntary actions in general, and not only of those that are involuntary due to factual error. I think that the category of the '*akôn*-agent' (from now on, 'the involuntary agent') can be easily extended[10] so as to cover *all* agents who feel (the relevant sort of) pain for having acted involuntarily, where 'pain' stands for a whole range of simultaneous and subsequent avoidance-attitudes like distress, reluctance, disgust, and of course, regret (*metameleia*). Likewise, the

---

[8] See Potts, T. C. (1965), p. 68.
[9] Taylor, C. C. W. (1965), p. 86-7.
[10] As Hursthouse, R. (1984) and Irwin, T. (1988) do.

category of the '*ouch hekôn-agent*' (from now on, 'the non-voluntary agent') could be extended to cover all those agents who do not feel (the relevant sort of) pain for having acted involuntarily. No doubt, the legitimacy of this extension can be disputed by the fact that such distinction between these two agents is never drawn in connection with violence (*bia*). It has been argued, for instance, that the reason "why Aristotle does not acknowledge the possibility of non-voluntary action [in my terminology, 'of the non-voluntary agent with regard to his involuntary action'] except in cases involving ignorance" is that "it is part of the definition of force that you cannot be forced to do something which is not contrary to your internal impulses, so there is no room for non-voluntary action due to force".[11] I have argued in Chapter 3 that this is not so (Section F): Aristotle recognizes, and is committed to recognize, a core definition of violence. Because of this, he is plainly committed to acknowledge a distinction analogous to the one between the two 'subcategories of the involuntary through factual error', or rather, analogous to the two categories of the agent's *status* with regard to his involuntary action, as I want to argue.

Given then that a piece of behaviour (i.e. 'what the agent did') is involuntary (genus) regardless of how the agent happens to feel about it, and keeping in mind Aristotle's suggestion that there is a further distinction between agents depending on their attitude towards this involuntary action, we can now formulate the pain condition (PC) as follows:

PC: For any piece of behaviour *B* that is performed involuntarily, an *agent* S is now involuntary (*akôn*) with regard to *B* if he now feels pain for causing or having caused *B*; and S is now *non-voluntary* (*ouch hekôn*) with regard to B if he now feels no pain for causing or having caused *B*.

Now, when I suggested at the beginning that the robust notion of violence should be understood as conjoining two *different* conditions, the pain condition and the externality condition, what I meant is this: (i) the externality condition (Aristotle's core notion of violence) and what we could now call 'the core notion of factual

---

[11] Whiting, J. E. (1988), p. 163-4. See also Means, T. (1927), p. 83-84.

error', tell us whether the piece of behaviour *B* is involuntary in the first place; that is, the significance of the core definition of violence and core definition of factual error is that they answer the demand for a sense in which *actions* are involuntary (genus); whereas (ii) the contrariety and pain conditions tell us whether *the agent* is now involuntary or non-voluntary with regard to *B* - the robust definition of violence misleadingly conjoins (i) and (ii).

The precise meaning of this principle, PC, won't become clear until Section D. Next, we must turn to an account of the sort of pain that defines an involuntary agent, according to Aristotle, but before we must polish our account of the pain condition.

B. Refining our account of PC

Until now I have merely been focusing on the pain condition, but we saw in Chapter 3 that the robust definition of violence also includes (besides the externality condition) a contrariety condition (i.e. 'against S's impulse (*hormê*)'). If the relation between the pain and contrariety conditions were merely that the former materially implies the latter (i.e. if there is pain, there is contrariety), this would be misleading, because it would allow, *at most*, for an understanding of pain as *evidence* of the corresponding contrariety to impulse. I think, however, that pain should be understood as a *truth condition* (or perhaps even as the *meaning*) of 'x is contrary to S's impulse'.

This is precisely the place to discuss this problem, because the fact that the pain involved in cases of factual error is contingent upon the agent's subsequently learning the relevant facts may suggest that the relation between pain and the contrariety condition could be *merely evidential* (pain is just a symptom of contrariety to impulse),[12] and thus 'x is/was against S's impulse' is not made true by S's subsequent regret, for it has been true all along. It is, however, very significant

---

[12] This is Sauvé-Meyer's view: "Pain and remorse, however, are only contingent results of contrariety to impulse, for an action can be contrary to an agent's impulse even if she never actually feels pain at or regret for it. This will be true, for example, of an agent who, because of ignorance of particulars, performs an action that is contrary to her desire, but never discovers what she has done. Therefore, as in the case of forced movement, pain is best understood as a symptom of the necessary contrariety rather than as a condition necessary in itself." Sauvé-Meyer, S. (1988), p. 133.

that Aristotle's 'robust conception' of factual error only involves the pain condition. For all Aristotle says in *EN* 1110b18-24, it is *the presence of regret* that makes an agent S involuntary (*akôn*) vis-à-vis a piece of behaviour *x* that S caused due to factual error, not the fact that *x* has been contrary to S's impulse all along. This of course, does not have to imply that there is no contrariety to impulse in cases of factual error where the agent subsequently regrets having caused *x* (in a rather trivial sense, there is), but it does imply, I think, that if there is some contrariety to impulse, such contrariety and the corresponding pain stand to each other at the very least in a relation of mutual implication. Accordingly, my suggestion in Chapter 3 (Section B) has been to take the pain condition, not only as implying contrariety to impulse, but also as *implied by* contrariety to impulse. Perhaps one could make the stronger claim that pain (adequately understood) is just what contrariety to impulse *means* ('stronger', that is, under the reasonable assumption that meaning cannot to be analysed merely in terms of truth conditions).

Still, that regret is contingent upon the agent's learning the relevant facts poses a problem for our formulation of PC, because it seems implausible to make the truth of 'S is an involuntary agent vis-à-vis a piece of behaviour B that is involuntary due to S's factual error' depend on the contingency of S's being lucky enough to learn the relevant facts. What if S is *such as* to feel regret, even if he does not for the contingent reason that he has not been lucky enough to learn the relevant facts? So perhaps PC should be partly rewritten in terms of subjunctive conditionals:

PC*: S is an *involuntary agent* towards an involuntary piece of behaviour *B*, if (i) S feels the relevant pain for causing *B*, or (ii) S would feel the relevant pain if S were to know the relevant facts; and S is a *non-voluntary agent* towards *B* if neither (i) nor (ii) are true of S.

PC* also has the advantage of taking care of 'mixed cases'. I have just taken for granted Aristotle's intuition that there is a temporal difference in the way pain accompanies forced motion (i.e. synchronically) and the way in which it accompanies factual error (i.e. subsequently, and on condition of the agent's acquiring the relevant piece of knowledge), but is it not also possible to conceive cases of violence

accompanied by *ex post facto* pain? To the extent to which such cases are conceivable, it is only because there is some sort of unawareness involved in them (i.e. because they are not 'pure' cases of violence). For example, suppose you are hypnotized by a crooked hypnotist into stealing jewels (as in Woody Allen's 'The Curse of the Jade Scorpion'), and that while acting under hypnosis you are fully aware of the relevant facts. In so far as the source of your action was external, you stole the jewels involuntarily through violence. Furthermore, your non-hypnotized self (let us suppose) is against the stealing of jewels, so that it is reasonable to think that you are an involuntary agent (*akôn*) with regard to the stealing of jewels. And yet, your discomfort, distress, etc. is subsequent to your stealing of jewels. But note: this distress is conditional upon your learning, once hypnosis is no longer operating, about what your hypnotized self did (indeed, Aristotle would probably say that this is a case of having but not using one's awareness, e.g. that these are someone else's jewels).

C. Moral pleasure and pain

It is not a mere accident that Aristotle opens the Nicomachean discussion on the voluntary, in the same Book that presents his account of the main ethical virtues (i.e. *EN* III), by asserting that "excellence has to do with affections and actions (*peri pathê te kai praxeis*)" and that "praise and blame are bestowed upon *the ones that are voluntary* (*epi tois hekousiois*)" (*EN* 1109b30-31); where '*epi tois hekousiois*' seems to refer back to *both pathê* and *praxeis*. But what does Aristotle mean by '*pathê*' here and in what sense can *pathê* be ´voluntary'?

To get hold of the relevant sense of '*pathê*' we need to examine the claim that "excellence has to do with feelings and actions (*peri pathê te kai praxeis*)" (*EN* 1109b30). This is a recurrent theme throughout the Aristotelian analysis of the nature of virtue and vice (*EN* 1104b14, 1106b16, b25, 1152b5-5; *EE* 1220a34). One of its most natural meanings, is that each virtue (perhaps with the sole exception of justice) has an specific domain of emotions within which it is expected to operate, by disposing the agent to be affected by it in the appropriate ways. Notice that there is nothing so inextricably obscure about the suggestion that *pathê* in this sense are

voluntary, in so far as this is taken to mean merely that we are praised and blamed in virtue of the appropriateness or inappropriateness of the emotions we have. I mean 'praised' or 'blamed' understood in terms of ethical ascription. For, as a matter of fact, we are called (by apt spectators) 'irascible' in virtue of our being angered too quickly and with too many people, 'cowardly' in virtue of our being afraid too easily, or 'envious' in virtue of our feeling pain at the sight of prosperity more often than we ought.

For our present purposes, there is, nonetheless, a more significant sense in which *pathê* figure among the concerns of ethical dispositions, namely, that ethical excellence has to do with *pleasures* and *pains* (*peri hêdonas kai lupas estin ê êthikê aretê*, *EN* 1104b9).[13] Pleasures and pains, in so far as they are the concern of ethical excellence (i.e. 'moral' pleasures and pains), are clearly among the *pathê* referred to in the opening line of *EN* III (1109b30). This is rendered plausible by Aristotle's claim that pain and pleasure (at any rate, the sort of pain and pleasure we are now concerned with) is a sign (*sêmeion*) of ethical dispositions (*EN* 1104b4), just as the particular ways of being affected are a sign of this.

Now, what is the distinctive mark of the pain and pleasure that is a sign of ethical dispositions? Interestingly enough, Aristotle indicates that it is the sort of pain and pleasure that is *supervenient on* (*epigignomenê, EN* 1104b5) or *follows upon* (*hepetai, EN* 1104b14, 1105b23)[14] the particular ways of being affected and acting that are peculiar to a given ethical disposition. *Every* characteristic emotion and *every* action is said to be followed by pleasure and pain. Incidentally, this turns out to be capital for purposes of moral education, for it is precisely because the particular range of feelings that constitute the field within which moral character operates (anger, fear, etc.) is characterized by being accompanied by pleasure and pain that they are amenable to habituation (*EE* 1222a6-14, 1227b5-10).

Narrowing down our initial claim that excellence is concerned with *pathê*, to the claim that it is concerned with the pain and pleasure that supervenes on, and follows upon, the emotions *and* actions characteristic of an ethical disposition (i.e.

---

[13] Also *EE* 1220a34-39, 1221b35-39, 1227b1-2, 1227b5-6; *MM* 1185b37.

[14] Also *EE* 1220b14; *Rhet.* 1378a21; *MM* 1186a13. The verb '*hepomai*' does not have to imply that pleasures and pains either *temporally follow* these feelings and actions, or that they are attendant upon them in the sense that they are attendant upon their actual *occurrence*.

the excellence's operations, its *erga*) is of course very helpful, but the meaning of this latter claim still remains quite obscure. In respect to particular *pathê* (i.e. the characteristic affections and emotions)*,* let us just note that the claim that these are 'accompanied' by pains and pleasures means, at the very least, that that they are "defined (*diôristai*) in terms of pleasure and pain" (*EE* 1221b37, cf. *Rhet*. 1378a21).[15] What is essential for my present purposes is that Aristotle does not restrict the claim that ethical dispositions are concerned with pleasures and pains, to the pleasures and pains accompanying the characteristic affections and emotions, but unambiguously extends this thesis to *praxeis*, actions (e.g. *EN* 1104b14). It is with this extension of moral pains and pleasures which are the concern of virtue, to the domain of action, that I am concerned in this section. Aristotle has already noted in Book I of the *EN* that "just acts are pleasant to the lover of justice and in general virtuous acts to the lover of virtue" (1099a10-11), and that "virtuous actions are in themselves pleasant" (1099a22), and argued in Book II that

The pleasure and pain that supervene on what people do (*erga*) should be treated as a sign of their dispositions; for someone who holds back from bodily pleasure and does so cheerfully (*chairôn*) is a moderate person (*sôphrôn*), while someone who is upset (*achthomenos*) at doing so is self-indulgent (*akolastos*), and someone who withstands frightening things and does so cheerfully, or anyway without distress (*mê lupoumenos*), is a courageous person, while someone who is distressed (*lupoumenos*) at them is cowardly. *EN* 1104b5-8[16]

It is worth noticing, however, that Aristotle explicitly recognizes that demanding that the virtuous agent positively rejoices in doing the virtuous action is far too stringent. (Although perhaps the verb '*chairô*' here used, indeed a verb usually contrasted with the feeling of pain in the discussion on the voluntary,[17] could be taken to mean simply 'welcoming' or 'accepting'). Aristotle recognizes that it is enough to demand that a person withstand frightening things or does something courageous *without distress* (*mê lupoumenos*) in order for him to be courageous

---

[15] On this see Leighton, S. R. (1996) and Frede, D. (1996).

[16] Rowe's translation. Aristotle indeed refers to this very same idea in his discussion of voluntariness: "those who act through violence and who are involuntary agents act with pain, but those who do acts for their pleasantness and nobility (*dia to hêdu kai kalon*) do them with pleasure (*meth' hêdonês*)" (*EN* 1110b11-13).

[17] *EE* 1224a31, a37, b17, b19, 1225a25, 1227b10.

(also *EN* 1154b24). So from now on I will talk about 'absence of pain' or better, 'compliance-attitudes' instead of 'pleasure'.

There is also an interesting ambiguity in the 1104b5-8 passage. Notice first that the person who is upset at holding back from bodily pleasures is here called '*akolastos*', 'self-indulgent': later on in *EN* VII Aristotle will define this sort of agent as '*enkratês*', 'continent', the self-indulgent being rather the 'perfectly' faulty person who pursues pleasure deliberately and without regret. Accordingly, if the contrast in 1104b5-8 supra is taken to be the one between *the continent* (rather than the self-indulgent) and the moderate person, perhaps the contrast between the cowardly and the brave *is not* the contrast between (i) the person who *does not* withstand frightening things because the prospects of doing so are too painful (i.e. the 'perfect' coward) and (ii) the person who does withstand frightening things and does this without distress; rather, the contrast seems to be between (i') someone who *does* withstand frightening things but does so *reluctantly*, and (ii') someone who withstands frightening things without distress. Now, the person who *does* withstand frightening things but does so reluctantly may not be a brave person, but he is certainly not a full-fledged coward either! So perhaps, following Grant (see his commentary *ad loc.*), one should say that Aristotle's point here is that 'self-indulgent' and 'cowardly' in this context are the contradictories of 'moderate' and 'courageous', the point being that (a) someone who is upset at holding back from bodily pleasure, but voluntarily does so anyway, is only non-moderate, and (b) someone who is upset at withstanding frightening things, but voluntarily does so anyway, is non-courageous. It is almost as if Aristotle was saying that (a) and (b) are 'involuntary agents' (*akontes*) vis-à-vis their voluntarily doing a moderate act and a courageous act respectively. This is important, for it will provide a model (in the last section) to understand what Aristotle means by 'involuntary agent' and 'non-voluntary agent' in PC*.

Now, some crucial qualifications must be added to the view just examined. For even the perfect exercise of courage will bring pain with it most of the time: "death and wounds will be painful (*lupêra*) to the courageous person and unwilling (*akonti*)" (*EN* 1117b7-8), and Aristotle goes even further by asserting that

the greater the extent to which he possesses excellence in its entirety, and the happier he is, the more he will be pained at the prospect of death; for to such a person, most of all, is living worthwhile, and this person will knowingly be depriving himself of goods of the greatest kind, which is something to be pained at. *EN* 1117b10-13[18]

This leads Aristotle to admit that "not all the excellences give rise to pleasant activity, except to the extent that pleasant activity touches on the end itself (*tou telous ephaptetai*)" (b15-16). This qualification on the view that virtuous activity is 'pleasant' is grounded on an important distinction between different *aspects* of the virtuous activity. Aristotle's solution consists in arguing that the courageous person is pleased, or at least not disturbed, by achieving the fine (i.e. the *telos*), and that this does not have to imply the absurdity that he is also pleased or undisturbed by getting wounded or sickly glad at the prospect of death (i.e. those things *by means of which* he may achieve the *telos*). Another way to express this is to say that the courageous person does what he does gladly or without pain under the description 'doing something admirable or fine' (i.e. the description of the action that 'touches on the end itself'), but not necessarily under the description 'dying' or 'getting wounded'.

So let the following qualification be added to our concept of 'moral' pain and pleasure or rather, moral aversion- and compliance-attitudes:

(I) They are the sort of attitudes that are *directed at* the ethically salient aspects[19] of an action or event.

This feature of moral aversion- and compliance-attitudes implies that these attitudes are not mere physiological, biologically-based pains and pleasures, but rather intentional states with a rich cognitive content. Moreover, it is particularly important, because it allows one to understand the pain and lack of pain in PC* (i.e. those attitudes that make an agent involuntary or non-voluntary vis-à-vis his *involuntary* action) also as moral aversion- and compliance-attitudes. This

---

[18] Rowe's translation.
[19] By the 'ethically salient aspect' of a piece of behaviour I mean a piece of behaviour done under a description that can ground an evaluative description in virtue- or vice-terms.

understanding is possible because *involuntary actions* (or motions) *can also have ethically salient aspects*, according to Aristotle. Aristotle recognizes that an involuntary action or state of affair can be described as 'something unjust' (*di adikon*) or as 'something moderate' (*ti sophrôn*), while recognizing that it is not thereby blameworthy or praiseworthy (*EN* 1135a9-11, 19-23; perhaps also implied in 1105b5-10).[20]

Interestingly enough, there is another feature of a moral aversion- or compliance-attitudes, now directed at the ethically salient features of *involuntary* behaviour, which indeed confirms their rich cognitive content. Suppose A strikes C using B's hand, and B had decided to strike C anyway. The externality condition is fulfilled, but not the contrariety condition, provided that B does not experience the relevant sort of pain *in striking C*. The problem is that B could not only feel pain *against striking C*, but also *against being forced by A*, notwithstanding his attitude towards the striking of C.[21] So another feature of the moral aversion- and compliance-attitudes directed at involuntary behaviour is this:

(II) Aversion- and compliance-attitudes (towards involuntary behaviour) are directed at the forced event (or *its* ethically salient features), rather than at the fact that the event is being forced. The same condition applies to cases of factual error, in which case the requirement is that these attitudes be directed at the action brought about through factual error (or *its* ethically salient features), rather than at the fact that the action had been brought about through factual error.

These features of moral-aversion and compliance attitudes, and in particular the ones directed at (the ethically salient aspects of) involuntary behaviour, strongly

---

[20] I am not making here – I think – the stronger claim that involuntary actions or states of affairs can be *good* or *bad* (on this see footnote 15, p. 70).

[21] It is important to note that the case in which the agent refuses to, and feels pain for, being compelled by violence full stop, is restricted in an important way: it is a case in which the agent *wants* in some sense to do *X* himself. The description of someone struggling and feeling pain *merely* for being compelled, when the thing he is compelled to do is something he does not want to do at all, seems to be only applicable to someone who is unaware of what he is being forced to do, otherwise it seems much more plausible to think that the agent is in pain for *being forced to do X*. This could explain why Aristotle seems to be unaware that pain can directly relate to the externality condition rather than to the contrariety condition, namely, because he focuses on cases in which the agent *does not* want to do X. Still, I want to suggest that Aristotle has good reasons to focus on the sort of pain that is connected to contrariety to impulse (in particular, connected to contrariety to reason, as I want to say).

suggest that they have a rich cognitive content. It is doubtful, for instance, that my dog is capable of holding an avoidance attitude towards my forcing him *to take a shower* as distinct from its *being forced* to take a shower. Indeed, taken together, these two features strongly suggest that the attitudes in question are based on the *rational* part of the soul. This is particularly important, because it confirms my interpretation of the kind of impulse (*hormê*) involved in the contrariety condition, as it figures both in the robust notion of violence (see Chapter 3, Section E). As I suggested in Chapter 3, if the connection between pain and the contrariety condition is as strong as I have argued, it is natural to think that one can arrive at the right interpretation of 'impulse' (*hormê*) in 'against S's impulse' by first determining the nature of the pain involved in the pain condition. The features of moral aversion-attitudes just mentioned allow us to do precisely this.

Consider again the case of a diabetic person who craves for a chocolate bar (a biologically based urge) and is compelled not to eat it by an external agent. His omission is involuntary, because it is the product of violence (according to the core notion). Now, suppose that, on the one hand, he feels the sort of pain associated with the thwarting of his appetite, and on the other he feels a 'rational' pleasure or lack of pain associated with the thought that the result is good for him overall. Why not think that the rational pleasure he harbours is indeed a moral compliance-attitude, in that it is directed (i) at the forced *event*, rather than the fact that the event is being forced, and (ii) directed at it under its ethically significant aspect - whereas the biological pain he also feels cannot possibly be a moral aversion-attitude? Let us grant that the diabetic agent in the example could not be temperate (unlike the unqualified continent, the temperate agent is not distressed by the non-satisfaction of a biologically-based craving). The pleasure and pain felt by the diabetic person rather exhibit his divided condition, his continent or incontinent condition. It is clear that the relevant attitude, for the purposes of pronouncing on the agent's ethical quality, is his 'rational' satisfaction or lack of pain concerning the overall results of the forced motion for his well-being, rather than his quasi-biological pain pointing at the opposite direction.

In general, since moral compliance- or aversion-attitudes involve a certain degree of capacity for discriminating the salient features of a practical situation (as

captured by (I) and (II)), and given that it is implausible to think that biologically-based cravings and non-rational impulses can ever perform this function, it is reasonable to think that the impulse relevant to the robust notions of violence and factual error (that is, the impulse contrariety to which makes an agent involuntary with regard to involuntary conduct) is his rational impulse (*boulêsis*) aiming at the good (*Rhet*. 1369a2-7) – as suggested in Chapter 3, Section F.[22] As we shall see in the next section, this has an important consequence, namely, that it does not exclude incontinent (nor continent) agents from the domain of agents who are capable of holding moral aversion- or compliance-attitudes towards a given piece of behaviour. Nor does it exclude children from agents who are so capable; indeed, it is very plausible to think that the gradual adoption of moral aversion- and compliance-attitudes is an essential ingredient in the process of moral education. Even though he does not say so explicitly, Aristotle is justified in claiming that only agents capable of holding compliance- and aversion-attitudes (thus defined) can be *voluntary agents*, and it is natural to think that he would identify these agents with *reason-responsive agents*, since any reason-responsive can in principle exercise the discriminatory capacities required by the holding of these moral attitudes.

D. The ethical significance of moral aversion- and compliance-attitudes

It may seem unobjectionable at first sight that the significance of there being moral aversion- and compliance-attitudes towards one's behaviour (whether voluntary or not) has to do most conspicuously with the fact that pain and pleasure are motivational forces, indeed the most primitive constituents of one's motivational set (*EN* 1104b30-1105a1). Let us first examine the relevance of this motivational aspect to the moral aversion-attitudes we are now considering, i.e. regret and the focused pain accompanying violent motion. One is certainly entitled to think that there is a close connection between motivation and, say, the moral aversion-attitude a subject takes towards an event brought about in his body through violence. It is

---

[22] Recall V2**, now rephrased as follows: S is an involuntary agent with regard to a piece of behaviour *B* which is involuntary through (core) violence, iff *B* is against S's rational impulse.

tempting to think that this is the sort of agent who *would not* have brought about the very same event in the absence of violence. And the same seems to be true about an agent who regrets an action due to factual error upon learning the true facts of the matter: it also seems to be the case that such a person is someone who would not have done the very same action voluntarily. Hence, it seems that one should conclude from the preeminence of the motivational aspect of aversion-attitudes that these derive their ethical significance solely from their being signs that the agent *would not* voluntarily do the action he is now distressed at causing or having caused involuntarily.[23] But is this really true after all? I don't think so. Let me offer two arguments in support of this negative answer.

First, consider a case of OR-coercion. The father who helps the criminal escape in order to save his daughter's life certainly does so reluctantly and in pain (or 'suffering pain', *EN* 1110a21), a pain focused merely on the incomplete description of the action as 'helping the criminal escape', a description of the situation to which, we can assume, the father inevitably relapses from time to time. And yet, helping the criminal escape is an action that the father does voluntarily, albeit under its complete, teleological description. The problem of course is that under that complete description, the action is not really an object of the agent's aversion-attitude, for the thought that he is saving his daughter's live is rather comforting. But this should not prevent us from concluding that, if two weeks after the incident the father helps a criminal escape involuntarily through factual error (say, believing falsely that he was a tourist asking for a lift) and subsequently regretted the outcome upon learning the truth, this regret *is not* necessarily a sign that he would *not* have helped a criminal escape voluntarily, because two weeks ago he did so in order to save his daughter! Perhaps then we should conclude that the agent's painful subsequent regret for having involuntarily caused *x* only shows that the agent would not have voluntarily done *x for its own sake* (*kath' auto*).

Now, this latter claim is very different from the conditional claim about what the agent would voluntarily do in a different scenario. If the pain in question corresponds indeed to a moral-aversion attitude, the primary message it sends – its

---

[23] This interpretation is not uncommon. See for instance Moline, J. N (1989).

motivational aspect notwithstanding – is that the action which is being imputed to the agent is such that it appears as something bad to the agent. That is, such that it appears as bad in virtue of the agent's ethical disposition.

Secondly, that the ethical significance of aversion-attitudes is not restricted to the conditional motivational truth about what the agent having those attitudes would do in a different scenario, also receives support from the nature of *metameleia*, which I have been translating as 'regret'. '*Metameleia*' means literally 'after-concern' or 'after-care'. The agent who feels *metameleia* for having caused a wrong action is someone who 'cares' or 'is concerned about' having caused it. I take it that his 'caring about' having caused a wrong action is not conditioned by the fact that the wrong action has been caused involuntarily. *Metameleia* is certainly an attitude one can have towards one's involuntary actions, but Aristotle also thinks that the incontinent is prone to regret his actions (*EN* 1150b29-1151a1) and he thinks that his incontinent behaviour is fully voluntary. What then is this 'caring about' having caused a blameworthy state of affairs? Well, it is certainly not the feeling of guilt that characterizes someone who thinks he *could have avoided* causing the wrong action*.* Presumably, *metameleia* is directed both at the agent's involuntarily causing a bad state of affairs through culpable error *as well as* his involuntarily causing it through an error for which he is not culpable at all. Nor is it plausible to think that there are two sorts of *metameleia*, one appropriate to voluntary actions and the other appropriate to involuntary ones. Now, although there is a motivational dimension to *metameleia*, the tendency to feel *metameleia* does not necessarily go with aversion that effectively motivates one to refrain. The incontinent agent feels regret for having done *x* against his rational considerations, but this does not stop him from voluntarily doing *x* over and over again. The case of the incontinent agent feeling *metameleia* suggests that this is the sort of feeling that speaks on behalf of the agent's 'deep self', that is, on behalf of the agent's own conception of the good (which, in the incontinent, remains intact), and it speaks thus when the agent is confronted with a piece of behaviour that is at variance with this evaluative conception, whether such piece of behaviour is voluntary or not. Accordingly, the primary message conveyed by *metameleia* (as it figures in PC) is that the action which is being imputed to the agent is such that it appears as

something bad to the agent, that is to say, bad in virtue of the agent's ethical condition.

Indeed, if the incontinent agent is allowed to hold a moral aversion attitude towards motion *M* or action *A*, then clearly his holding this aversion attitude towards *M* or *A* is not a sign that he would *not* voluntarily do *M* or *A*, for his being incontinent precisely consists in his doing *M* or *A* voluntarily despite his holding the aversion attitudes against *M* or *A*.

Notice also that the motivational aspect of moral aversion- and compliance-attitudes directed at involuntary actions (captured by the conditional claim about 'what the agent would do') vanishes as we turn to moral *compliance*-attitudes. This is not only because compliance-attitudes are not necessarily aligned with positive pleasures, for even if they were so aligned, the fact that an agent takes 'moral' delight on an event involuntarily brought about is clearly not a sign that he would bring it about voluntarily in the absence of violence or factual error. If A is pushed by a gust and pushes B as a result, B falls into a ditch and gets badly injured, and A is not distressed to see B injured (if you want, imagine B is rather glad for the outcome); why should we think that A would have pushed B into a ditch if he had had the chance to do so? And yet, if B's lack of distress is the relevant one, that is, if it corresponds indeed to a moral compliance-attitude, then it tells us something about B's evaluative outlook on A's accident.[24]

E. The compatibility test

Recall the way in which I have formulated, in Chapter 2, the question about Aristotle's concerns with the voluntary in his Ethics:

---

[24] It seems to me that one can see the significance of this identification or disassociation attached to moral compliance- and aversion-attitudes more also in the case of those actions performed by the other members of the agent's community. Although Aristotle does not underscore the importance of this other-regarding aspect of the attitudes in question, it is clear that one can identify oneself with any of the ethical actions performed by members of one's moral community by means of approving to them, satisfaction at their being done, or perhaps simply by being pained at their being done, just as one can dissociate one self from them by means of one's aversion-attitudes towards them. One can then see the agent's aversion- or compliance-attitudes towards those events involuntarily caused by himself as analogous to his attitudes towards actions performed by members of his own ethical community, and perhaps see both self-addressed and other-addressed attitudes as part of a general category of attitudes towards ethically significant events that are not voluntarily caused by the attitude-holder himself.

Qe: What are the conditions that need to be fulfilled in order for an apt observer to be warranted in (1) praising and blaming a piece of behaviour B as 'v' (on the basis of criterion C) and (2) praising or blaming a reason-responsive agent S in 'V'-terms through (*dia*) (1)?

When engaged in ethical ascription we (the virtuous spectators) naturally start from an observed piece of behaviour that is assumed to be voluntary and praiseworthy or blameworthy (i.e. we usually assume (1)). We see the warrior fleeing from the battle field, and assume this to be a voluntary, cowardly action. Notice that the cases were this assumption is not made in the first place are cases in which we know the agent beforehand, and trust he is not the sort of person to act cowardly. On this belief we proceed to see if the agent did the action in full knowledge of his circumstances, whether he was forced to do it or not, whether he acted on an overwhelming impulse, or whether he was objectively justified. If none of these rebutting explanations is available, and his fleeing is voluntary, we conclude that the agent is not brave as we thought. Now we conclude he is not brave, but are we allowed to call him a cowardly person (i.e. to move on to step (2))? On Aristotle's view, we are not, unless we have established that he also displays the relevant compliance-attitude *towards the cowardly action as such*. For all we know, he could be an incontinent agent, and he could have left the battle-field because he had to attend to a game of poker (although this may ground our calling him 'irresponsible', etc. it does not ground our calling him 'cowardly').

But now, if any of the aforementioned rebutting explanations is available at level (1), and the agent has acted involuntarily, this is not the end of the matter, from the point of view of ethical ascription. This is precisely the consequence of PC*. Aristotle's point is that S's moral aversion- or compliance-attitudes tell us something about the agent's evaluative outlook, his conception of the good, even though the piece of behavior towards which these attitudes are directed is involuntary (or perhaps even a piece of behavior or event caused by some other agent), and even though they do not necessarily tell us what the agent would voluntarily do in a counterfactual scenario (for this would just mean that aversion- and compliance-

attitudes are in the end just attitudes towards voluntary actions, rather in the way motivations are). On my interpretation, aversion- and compliance-attitudes can be attitudes towards involuntary pieces of behavior all the same, because even involuntary pieces of behavior can have ethically salient features. We have seen that Aristotle recognizes this (p. 196-197).

I am now in a position to maintain that, if a putative agent has a moral aversion-attitude towards an involuntary piece of behaviour B and its ethically salient features, e.g. as something unjust (*ti adikon*), this of course defeats the assumption of voluntariness at level (1) in the most successful way possible: namely, by signaling the fact that he is a non-unjust agent: this is the force of '*involuntary agent*'. A compliance moral attitude towards B as *ti adikon*, on the other hand, affects in a positive way the ethical status of the agent: we are not justified in calling him 'unjust', but we know *he is* a *non*-just agent (and for all we know he *could* be unjust): this is the force of 'non-voluntary agent'. The force of 'involuntary agent' and 'non-voluntary agent' is explained by Aristotle's focus on the involuntary causing of wrong actions, but the distinction also applies when the salient features of a piece of behaviour B are positive (though not 'praiseworthy'), e.g. 'something moderate (*ti sôphron*).' Here a moral aversion-attitude towards B tells us something more positive about the agent: although we can't call him 'immoderate', we know *he is* a non-moderate agent. If his attitude is a moral attitude of compliance, although we can't call him 'moderate', we know he is non-immoderate (for all we know he could be moderate). More succinctly put: aversion-attitudes are incompatible with the trait of character corresponding to the ethically salient features of B, whereas compliance attitudes are compatible (as our interpretation of *EN* 1104b5-8 in Section C suggested).[25] There is then a precise way in which the pain condition directly affects the status of the agent, that is, without the mediation (*dia*) of the agent's voluntary behaviour.

---

[25] A diagram would look like this:

Assumed character → Moral attitudes ↓	**Virtuous character** (on the basis of e.g. *ti dikaion*)	**Faulty character** (on the basis of e.g. *ti adikon*)
Aversion attitude	Incompatible (e.g. non-just agent)	Incompatible (e.g. non-unjust agent)
Compliance attitude	Compatible (e.g. non-unjust agent)	Compatible (e.g. non-just agent)

I take it that doubts that may arise as to why exactly should the warrior's reluctance and distress at his being dragged away from the battle field be *incompatible* with his being a coward, are the more acute the more unspecified the reluctance and distress in question remain. We have seen until now two main ways in which this can be so (corresponding to the two features of moral attitudes distinguished in Section C): (i) it can remain unspecified in the sense that we don't know whether the warrior's distress is distress at his being dragged away *from the battle field*, or at his *being dragged away*, full stop. Of course, we are particularly interested in the first sort of distress, but this sense itself can also remain unspecified, namely (ii) in the sense that we don't know whether it is distress at his being dragged away from the battle field *in so far as* this is something bad and cowardly, or *in so far as* the experience at the battle field involves things the warrior enjoys doing (e.g. killing people, hearing the beat of the drums, etc.). Again, the relevant distress here is the first one. The initial doubt then turns out to be: 'why should we think that the warrior's reluctance and distress at his being dragged away *from the battle field*, *in so far as* leaving the battlefield in this occasion is seen by the warrior *as something cowardly*, is incompatible with his being a cowardly person?' Thus expressed, the answer is virtually contained in the question.

APPENDIX

Two lexical ambiguities in Aristotle's use of '*eph' hautô*[*i*]'

In this appendix I attempt to clarify two lexical ambiguities involved in the notion of *to eph' hautô*[i] or 'what depends on one' (contingency power) as used by Aristotle in his Ethics.[1]

I. The first ambiguity has to do with the scope of 'it depends on S':

(1) Φ depends on S = It depends on S to Φ & it depends on S not to Φ

(2) Φ depends on S = It depends on S (to Φ & not to Φ)

I want to show first that (1) and (2) are unsatisfactory formulations of the notion of *eph' hautô*[*i*]. (1) seems unsatisfactory because it suggests that there are *two different* powers at stake, one for Φ-ing and one for not Φ-ing, that, although they may be mutually dependant, are indeed different. This is misleading because Aristotle clearly believed that *the same* contingency-power enabled one both to Φ and not to Φ. He thought that if I indeed possess at *t* the contingency power to type a sentence, then both when I abstain from typing a sentence at *t* as well as when I am typing a sentence at *t*, I am exercising or manifesting the *very same* contingency power.[2] Formulation (2), on the other hand, suggests that there is only one contingency power at stake. The problem with (2) is that it is misleading because it suggests that there is one single contingency power that can be actualized in these two ways *at the same time*, which is impossible (see *Met*. 1048a8-15). What generates this problem is Aristotle's frequent use of the conjunction '*kai*' (e.g. *EN* 1110a16). In what follows, I will use the disjunction '*mê*' (e.g. *EE* 1223a7-8):

(3) Φ depends on S = It depends on S (to Φ V not to Φ).

---

[1] E.g. *hosa d' eph' hautô*[*i*] *esti poiein ê mê poiein* (*EE* 1223a7-8); *ep' autô*[*i*] *kai to prattein kai mê* (*EN* 1110a16); *en hois gar eph' hêmin to prattein, kai mê prattein* (*EN* 1113b7).
[2] This at least is clear from his discussion of rational capacities at *Met*. IX 2 and 5. I am not of course implying that rational capacities and voluntary capacities are the same, but they are clearly similar in so far as neither of them involve two different capacities.

I want to argue that (3) is indeed the correct formulation of Aristotle's notion of *ep' hautô*[*i*].

II. In order to show this, still another ambiguity must be clarified: there is another sense in which the notion of *eph' hautô*[i] can be differently analysed, depending on the scope of the negation ('*mê*'):

(3) Φ depends on S = It depends on S (to Φ V *not* Φ)

(4) Φ depends on S = It depends on S (to Φ V to *not*-Φ)

'to *not*-Φ' and 'to Φ' in (4) are contraries, that is to say, 'to *not*-Φ' refers to a positive action contrary to Φ (call it 'Ψ').

Now, in cases in which there is only *one* action to be considered, Aristotle unambiguously formulates the notion of *eph' hautô*[i] in terms of (3). This may be obscured by Aristotle's elliptical formulations like '*ep' hautô*[i] *kai to prattein kai mê*' (*EN* 1110a16), or '*eph' hautô*[i] *taut' esti ginesthai kai mê*' (*EE* 1223a8), but it is clear that these formulae are elliptical for '*mê prattein*' and '*mê ginesthai*' respectively, as made clear by the definition of the voluntary at *EE* 1225b8 (see also 1225a10).

Formulation (4), on the other hand, may seem important because, in claiming that virtue and vice are voluntarily acquired dispositions, Aristotle wants to say that whenever it depends on me to do a virtuous action, it also depends on me to do a faulty one, and *vice versa* (for instance, *EN* 1113b6-7 and *EE* 1228a7-8), virtue- and vice-terms being contraries. So it is tempting to think that when Aristotle employs the notion of *eph' hautô*[i] in connection with the Platonic question about the voluntariness of virtue and vice, he uses a formula of the (4) type above. But in fact, Aristotle never says this explicitly. On the contrary, I want to show that in these contexts, it is a notion of *eph' hautô*[i] captured by (3), and not by (4), what Aristotle uses and needs.

EN III 5 argues that (sentences are numbered for easy of reference):

(*P*) when acting (*to prattein*) depends on us, not acting (*to mê prattein*) does so too, and when saying no does so, saying yes does too; so that (i) if acting, when it is a fine thing to act, depends on us, not acting also depends on us when it is shameful not to act, and (ii) if not acting, when it is a fine thing not to act, depends on us, also acting when it is a shameful thing to act also depends on us. But if it depends on us to do fine things and shameful things, and similarly not to do them too, and this, it is agreed, is what it is to be, respectively, a good person and a bad one, then being decent people, and being worthless ones, will depend on us.[3] *EN* 1113b6-14

The argument in the *EN* 1113b6-14 passage only seems to work under the assumption that Aristotle is working with a notion of *eph' hautô*[i] as captured by (3).

To show this, let me first rule out one possible interpretation of the passage. It is tempting to think that Aristotle is assuming in the *EN* 1113b6-14 passage a principle like:

P: It depends on us to Φ ↔ It depends on us *not* to Ψ

where 'Φ' and 'Ψ' are contraries (for instance, 'to do what is good' and 'to do what is bad'). Moreover, this is not an isolated passage, for *EE* 1228a5-7 repeats a similar argument, also in a similar context: "If a man, having it in his power to do the honourable and abstain from doing the base, does the opposite, it is clear that this man is not good".[4] Here the power referred to is *not* the power to Φ and to not-Φ (or not to Φ), but rather the power to Φ and *not* to (*apraktein*) Ψ (for instance, the power to do what is good and *not* to do what is bad), 'Φ' and 'Ψ' being contraries.

Now, one might wonder why Aristotle thought that this particular application of P – if P is indeed his assumption – is sufficient to establish that 'both virtue and, in the same way, vice depend on us" (*EN* 1113b6-7; *EE* 1228a7-8). As shown by the corresponding passage of the *Magna Moralia* (1187a5-8), Aristotle is here arguing against the Socratic Asymmetry Thesis, according to which 'no one does wrong voluntarily'.[5] Even assuming as Aristotle now does for dialectical purposes, that 'to *be* good/bad' and 'to *do* good/bad actions' are equivalent (see *EN* 1113b13) as Plato

---

[3] Rowe's translation.
[4] Salomon's translation.
[5] See Plato's *Protagoras* 345d-e; *Timaeus* 87b; *Laws* 860d.

sometimes suggests[6], it is clear that (i) and (ii) in the *EN* 1113b6-14 passage supra do not, taken separately, amount to arguments against the Asymmetry Thesis. Perhaps then (i) and (ii) should be taken together as an argument (the 'Logical Argument for the Symmetry Thesis') based on P, in which case, the argument would be something like this:

1. It depends on S to Φ → It depends on S *not* to Ψ

2. It depends on S *not* to Φ → It depends on S to Ψ

3. *Ergo*: It depends on S to Φ → It depends on S to Ψ

where 'Φ' stands for 'to do something good' and 'Ψ' for 'to do something bad'. The problem is that the argument is clearly assuming a further premise, linking the consequent of 1 with the antecedent of 2, but such a premise is not available merely on the basis of P.

To make the argument work, all Aristotle needs is (3) and a simple argument to show that the Asymmetry Thesis is false:

1. It depends on S to Φ → It depends on S *not* to Φ   (3)

2. It depends on S *not* to Φ → It depends on S to Ψ

3. *Ergo*: It depends on S to Φ → It depends on S to Ψ

Premise 1 depends on (3). Premise 2 does not have to rely on P and P's implausible assumption of a binary system with no *tertium quid* (e.g. doing what is required is good and not doing what is required is bad). Aristotelian virtue- and vice-terms do not in general behave binarily. All Aristotle needs in order to show the falsity of the Asymmetry Thesis is to offer *one* convincing illustration of premise 2; say, if it depends on me *not* to return a deposit and thereby *not* to do something just, it also depends on me to *do* something *un*just, precisely by not returning the deposit. I conclude then that all Aristotle needs, as a definition of *eph' hautô*[i] is formulation (3).

---

[6] The *locus classicus* is *Laws* IX, 860d. On this see Sauvé-Meyer, S. (1993), pp. 131-2.

Bibliography

Ackrill, J. L. (1978): 'Aristotle on action'. In Rorty, A. O. (Ed.): 1980, pp. 93-102. (Reprinted from *Mind* 87, 595-601)

Adams, R. M. (1985): 'Involuntary sins'. *Philosophical Review* 94, 3-31

Adkins, A. W. H. (1960): *Merit and Responsibility: A Study in Greek Values*. Oxford: Clarendon Press

Allan, D. J. (1936): *Aristotelis De Caelo*. Oxford: Clarendon Press

Annas, J. (1980): 'Aristotle on pleasure and goodness'. In Rorty, A. O. (Ed.): 1980, pp. 285-300

- (1992): 'Ancient ethics and modern morality'. *Philosophical Perspectives* 6, 119-136

Anscombe, G. E. M. (1958): 'Modern moral philosophy'. *Philosophy* 33, 1-19

- (1963): 'Two kinds of error in action'. *Journal of Philosophy* 60, 393-401

Gilby, T. (2006): *Thomas Aquinas, Summa Theologiae*, *Volume 17*: *Psychology of Human Acts*. New York: Cambridge University Press

Araujo, M. & Marias. J. (1970): *Aristóteles: Ética a Nicómaco*. Madrid: Instituto de Estudios Públicos

Austin, J. L. (1956-7): 'A plea for excuses'. In Morris, H. (Ed.): 1961, pp. 6-19. (Reprinted from *Proceedings of the Aristotelian Society* 57, 1-30)

Austin, J. (1873): 'Negligence, heedlessness, and rashness'. In Morris, H. (Ed.): 1961, pp. 240-246. (Reprinted from his *Lectures on Jurisprudence* (pp. 440-444). London: John Murray)

Baier, K. (1972): 'Types of responsibility.' In French, P. (Ed.): 1991, pp. 117-122. (Reprinted from French, P. (Ed.), *Individual and Collective Responsibility* (pp. 56-61). Cambridge, MA: Schenkman Publishing)

Baker, M. B. (1983): 'The excuse of accident.' *Ethics* 93, 695-708

Barnes, J. *et al*. (Eds.) (1977): *Articles on Aristotle, Volume 2*. London: Duckworth

Bennett, J. (1980): 'Accountability'. In Van Ztraaten, Z. (Ed), *Philosophical Subjects: essays presented to P. F. Strawson* (pp. 14-47). Oxford: Clarendon Press

Bodnár, I. M. (1997): 'Movers and elemental motions in Aristotle'. *Oxford Studies in Ancient Philosophy* 15, 81-117

Bondeson, W. (1974): 'Aristotle on responsibility for one's character and the possibility of character change'. *Phronesis* 19, 59-65

Bonitz, H. (1870): *Index Aristotelicus*. Berlin: George Reimer

Brandt, R. B. (1958): 'Blameworthiness and obligation'. In French, P. (Ed.): 1991, pp. 90-109. (Reprinted from Melden, A. I. (Ed.), *Essays in Moral Philosophy* (pp. 3-39). Seattle: University of Washington Press)

- (1970): 'Traits of character: a conceptual analysis'. *American Philosophical Quarterly* 7, 23-37

Bravo, F. (2006): 'Teoría aristotélica de la responsabilidad'. *Estudios de Filosofía* 34, 109-132

Brickhouse, T. C. (1991): 'Roberts on responsibility for action and character in the *NE*'. *Ancient Philosophy* 11, 137-148

- (2003): 'Does Aristotle have a consistent account of vice?' *The Review of Metaphysics* 57, 3-23

Broadie, S. (1991): *Ethics with Aristotle*. New York: Oxford University Press

- (1994): 'Another problem of *akrasia*'. *International Journal of Philosophical Studies* 2, 229-242

- & Rowe, C. (2002): *Aristotle: Nicomachean Ethics*. Oxford, New York: Oxford University Press

Burnet, J. (1900): *The Ethics of Aristotle.* London: Methuen

- (1905-1913): *Platonis Opera, Volumes I-V*. Oxford: Oxford University Press

Burnyeat, M. F. (1980): 'Aristotle on learning to be good'. In Rorty, A. O. (Ed.): 1980, pp. 69-92

Bywater, I. (1894): *Aristotelis Ethica Nicomachea.* Oxford: Clarendon Press

Charles, D. (1984): *Aristotle's Philosophy of Action.* Ithaca: Cornell University Press

- (2009): 'Nicomachean Ethics VII. 3: Varieties of *akrasia*'. In Natali, C. (Ed.), *Aristotle*: *Nicomachean Ethics*, *Book VII Symposium Aristotelicum* (pp. 41-71). Oxford, New York: Oxford University Press

Cohen, S. M. (1994): 'Aristotle on elemental motion'. *Phronesis* 39, 150-159

Cooper, J. M. (1975): *Reason and Human Good in Aristotle*. Cambridge: Harvard University Press

- (1973): 'The *Magna Moralia* and Aristotle's moral philosophy'. *American Journal of Philology* 94, 327-349

Curren, R. (1989): 'The contribution of Nicomachean Ethics iii 5 to Aristotle's theory of responsibility'. *History of Philosophy Quarterly* 6, 261-277

Darwall, S. (2006): *The Second-Person Standpoint: Morality, Respect, and Accountability.* Cambridge, London: Harvard University Press

Davidson, D. (2001): *Essays on Actions and Events* (2nd ed.). Oxford: Clarendon Press

Déclarie, V. & Houdé-Sauvé, R. (1978): *Aristotle*: *Éthique à Eudéme*. Paris: Librairie Philosophique J. VRIN

Di Muzio, G. (2000): 'Aristotle on improving one's character'. *Phronesis* 45, 205-219

Dover, K. J. (1994): *Greek popular morality in the time of Plato and Aristotle*. Indianapolis: Hackett

Duff, R. A. (1990): *Intention, Agency & Criminal Liability: Philosophy of Action and the Criminal Law*. Oxford, Cambridge (MA): Basil Blackwell

Dyer, R. R. (1965): 'Aristotle's categories of involuntary torts (E. N. v. 1136b8-25).' *Classical Review* 129, 250-252

Ellebaudt, N. (Ed.) (1565): *Nemesii Episcopi et De Natura Hominis, Lib. Unus.* Antuerpiae: ex officina Christophori Plantini

Everson, S. (1990): 'Aristotle's compatibilism in the *Nicomachean Ethics*'. *Ancient Philosophy* 10, 81-99

Feinberg, J. (1970): *Doing & Deserving*. Princeton: Princeton University Press

- (1986): *Harm to Self*. New York, Oxford: Oxford University Press

Fischer, J. M. & Ravizza, M. (1998): *Responsibility and Control*: *A Theory of Moral Responsibility*. Cambridge: Cambridge University Press

Foot, P. (1978): *Virtues and Vices and Other Essays in Moral Philosophy.* Berkeley: University of California Press

Fortenbaugh, W. W. (1975): *Aristotle on Emotion*. London: Duckworth

Fowler, M. (1982): 'Coercion and practical reason'. *Social Theory and Practice* 8, 329-355

Frankfurt, H. (1971): 'Freedom of the will and the concept of a person'. *Journal of Philosophy* 68, 5-20

- (1973): 'Coercion and moral responsibility'. In Honderich, T. (Ed.), *Essays on Freedom of Action* (pp. 65-86). London: Routledge & Kegan Paul

Frede, D. (1986): 'Mixed feelings in Aristotle's rhetoric'. In Rorty, A. O. (Ed.): 1996, pp. 258-285

- (2006): 'Pleasure and pain in Aristotle's ethics'. In Kraut, R. (Ed.): 2006, pp. 255-275

French, P. (1991): *The Spectrum of Responsibility.* New York: St. Martin's Press

Freeland, C. (1986): 'Aristotle on possibilities and capacities'. *Ancient Philosophy* 6, 69-89

Freese, J. H. (2000): *Aristotle, the Art of Rhetoric.* Cambridge, Massachusetts: Harvard University Press

Furley, D. J. (1967): *Two Studies in the Greek Atomists.* Princeton, N. J.: Princeton University Press

- (1977): 'Aristotle on the voluntary'. In Barnes, J. *et al*. (Eds.): 1977, pp. 47-60

- (1978): 'Self-Movers'. In Rorty, A. O. (Ed.): 1980, pp. 55-68. (Reprinted from Loyd, G. E. R. & Owen, G. E. L. (Eds.), *Aristotle on Mind and the Senses* (pp. 165-179). Cambridge: Cambridge University Press)

Gagarin, M. (1981). *Drakon and Early Athenian Homicide Law*. New Heaven, London: Yale University Press

Gauthier, R. A. & Joliff, J. Y. (1959): *Aristote: l'Ethique à Nicomaque* (2nd ed., 2vols). Louvain: Publications Universitaires

Gill, M. L. (1991): 'Aristotle on Self-Motion'. In Judson, L. (Ed.), *Aristotle's Physics*: *A Collection of Essays* (pp. 243-265). New York, Oxford: Clarendon Press

Grant, A. (1885): *The Ethics of Aristotle* (4th ed., 2vols.). London: Longman

Greenbaum, S. *et al.* (1972): *A Grammar of Contemporary English.* London: Longman

Grönroos, G. (2007): 'Listening to reason in Aristotle's moral psychology'. *Oxford Studies on Ancient Philosophy* 32, 251-272

Haksar, V. (1964): 'Aristotle and the punishment of psychopaths'. *Philosophy* 39, 323-340

Hardie, W. F. R. (1968): 'Aristotle and the free will problem [Reply to Huby 1967].' *Philosophy* 43, 274-8

- (1980): *Aristotle's Ethical Theory* (2nd ed.). Oxford: Clarendon Press

Hare, R. M. (1952): *The Language of Morals*. Oxford: Clarendon Press

Hart, H. L. A. (1949): 'The ascription of responsibility and rights'. In Morris, H. (Ed.): 1961, pp. 143-157. (Reprinted from *Proceedings of the Aristotelian Society* 59, 171-19)

- & Honoré, T. (1985): *Causation in the Law* (2nd ed.). Oxford, New York: Oxford University Press

Heinaman, R. (1986): 'The Eudemian Ethics on knowledge and voluntary action'. *Phronesis* 31, 128-147

- (1988): 'Compulsion and voluntary action in the Eudemian Ethics'. *Noûs* 22, 253-281

Heybut, G. (Ed.) (1889): *Aspasius in Ethica Nicomachea Quae Supersunt Commentaria*, *Volume 1*. Berlin: George Reimer

Hicks, R. D. (1925): *Diogenes Laertius*: *Lives of eminent philosophers*, *Volume 1*. London: Heinemann

Hobart, R. E. (1934): 'Free will as involving determination and inconceivable without it'. *Mind* 43, 1-27

Huby, P. (1967): 'The first discovery of the freewill problem.' *Philosophy* 42, 353-62

Hursthouse, R. (1984): 'Acting and feeling in character: Nicomachean Ethics 3.i.' *Phronesis*, 29, 252-66

Hutchinson, D. S. (1986): *The Virtues of Aristotle.* London, New York: Routledge & Kegan Paul.

Irwin, T. H. (1980): 'Reason and responsibility in Aristotle'. In Rorty, A. O. (Ed.): 1980, pp. 117-156.

- (1988): *Aristotle's First Principles.* Oxford: Clarendon Press

- (1996): 'Ethics in the *Rhetoric* and in the Ethics'. Rorty, A. O. (Ed.): 1996, pp. 142-174

- (2001): 'Vice and reason'. *The Journal of Ethics* 5, 73-97

- (2007): *The Development of Ethics, Volume 1*. Oxford, New York: Oxford University Press

Jaeger, W. (1957): *Aristotelis Metaphysica.* Oxford: Clarendon Press

Jackson, H. (1901): 'On Nicomachean Ethics III i §17, 1111a8, and Republic VIII 563 C'. *Journal of Philology* 27, 159-160

Joachim, H. H. (1951): *Aristotle, The Nicomachean Ethics*. Oxford: Clarendon Press

Johnson, K. (1997): 'Luck and good fortune in the *Eudemian Ethics*'. *Ancient Philosophy* 17, 85-102

Kassel, R. (1965): *Aristotelis De Arte Poetica Liber*. Oxford: Clarendon Press

Kenny, A. (1978): *Freewill and Responsibility*. London: Routledge & Kegan Paul

- (1979): *Aristotle's Theory of the Will*. London: Duckworth

Klimchuk, D. (2002): 'Aristotle on necessity and voluntariness'. *History of Philosophy Quarterly* 19, 1-19

Konstan, D. (2006): *The Emotions of the Ancient Greeks: Studies in Aristotle and Classical Literature.* Toronto, Buffalo, London: University of Toronto Press

Kosman, L. A. (1980): 'Being properly affected: virtues and feelings in Aristotle's ethics'. In Rorty, A. O. (Ed.): 1980, pp. 103-116

Kraut, R. (Ed.) (2006): *The Blackwell Guide to Aristotle's Nicomachean Ethics.* Malden, MA, Oxford: Blackwell Publishing

Lee. H. D. P. (1937): 'The legal background of two passages in the *Nichomachean Ethics*'. *Classical Quarterly* 31, 129-40

Leighton, S. R. (1996): 'Aristotle on the emotions'. In Rorty, A. O. (Ed.): 1996, pp. 206-237

Long, H. S. (1964): *Diogenis Laertii Vitae Philosophorum*. Oxford: Oxford University Press

Loomis, W. T. (1972): 'The Nature of Premeditation in Athenian Homicide Law'. *Journal of Hellenic Studies* 92, 86-95

Mackenzie, M. M. (1981): *Plato on Punishment*. Berkeley, Los Angeles, London: University of California Press

MacDowell, D. M. (1966): *Athenian Homicide Law* (2nd ed.). Manchester: Manchester University Press

- (1991): *Gorgias: Encomiun of Helen* (2nd ed.). Bristol: Bristol Classical Press

McDowell, J. (1979): 'Virtue and reason'. *Monist* 62, 331-50

Means, T. (1927): 'Aristotle and the voluntary'. *Transactions and Proceedings of the American Philological Association* 58, 75-91

Mele, A. (1984): 'Aristotle's wish'. *Journal of the History of Philosophy* 22, 139-156

- (1992): *Springs of Action: Understanding intentional behaviour.* New York, Oxford: Oxford University Press

Mingay, J. M. & Walzer, R. R. (1991): *Aristotelis Ethica Eudemia*. Oxford: Clarendon Press

Minio-Paluello, L. (1952): *Aristotelis Categoriae*. Oxford: Clarendon Press

Moline, J. N. (1989): 'Aristotle on praise and blame'. *Archiv für Geschichte der Philosophie* 71, 283-302

Morris, H. (Ed.) (1961): *Freedom and Responsibility*. Stanford, California: Stanford University Press

Murray, M. J., & Drudrick, D. F. (1995): 'Are coerced acts free?' *American Philosophical Quarterly* 32, 109-123.

Nagel, T. (1993): 'Moral luck'. In Statman, D. (Ed.): 1993, pp. 57-72

Natali, C. (2004): *L'Action Efficace: Études sur la Philosophie de l'Action d'Aristote.* Louvain: Éditions Peeters

Nauck, A. (1926): *Tragicorum Graecorum Fragmenta* (2nd ed.). Lipsiae: In aedibus B.G. Teuberneri

Nielsen, K. M. (2007): 'Dirtying Aristotle's hands? Aristotle's analysis of 'mixed acts' in the *Nicomachean Ethics* III, 1.' *Phronesi*s 52, 270-300

Nozick, R. (1969): 'Coercion'. In Patrick, S., & Morton, W. (Eds.), *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel* (pp. 440-472). Sidney Morgenbesser, New York: St. Martin's Press

Nussbaum, M. C. (1978): *Aristotle's De Motu Animalium.* Princeton: Princeton University Press

- (1986): *The Fragility of Goodness*. Cambridge: Cambridge University Press

Oshama, M. A. L. (1997): 'Ascriptions of responsibility'. *American Philosophical Quarterly* 34, 71-83

Ott, W. (2000): 'A troublesome passage in Aristotle's Nicomachean Ethics iii 5'. *Ancient Philosophy* 20*,* 99-108

Owen, G. E. L. (1960): 'Logic and metaphysics in some earlier works of Aristotle'. In Düring, I., & Owen, G. E. L (Eds.), *Aristotle and Plato in the Mid-Fourth Century* (pp. 163-190). Göteborg: Studia Graeca et Latina Gothoburgesia XI

Potts, T. C., & Taylor, C. C. (1965): 'States, Activities and Performances'. *Proceedings of the Aristotelian Society* (Supplement) 39, 65-102

Rackham, H. (1938): *Aristotle: The Athenian Constitution, The Eudemian Ethics, On Virtues and Vices.* Cambridge, Massachusetts: Harvard University Press

Roberts, J. (1989): 'Aristotle on responsibility for action and character'. *Ancient Philosophy* 9, 23-36

Richards, A., & Stratchey, J. (1991): *Sigmund Freud: New Introductory Lectures on Psychoanalysis, Volume 2*. London, New York: Penguin

Roemer, A. (1922): *Aristotlelis Ars Rhetorica*. Leipzig: Teubner

Rorty, A. O. (Ed.) (1980): *Essays on Aristotle's Ethics*. Berkeley, Los Angeles, London: University of California Press

- (Ed.) (1996): *Essays on Aristotle's Rhetorics*. Berkeley, Los Angeles, London: University of California Press

Rosen, G. (2003): 'Culpability and ignorance'. *Meeting of the Aristotelian Society* 103, 61-84

Ryle, G. (1949): *The Concept of Mind*. Harmondsworth: Penguin

Ross, D. W. (1923): *Aristotle*. London: Methuen's Publications

- (1956): *Aristotelis De Anima*. Oxford: Clarendon Press

- (1957): *Aristotelis Politica*. Oxford: Clarendon Press

- (1959): *Aristotelis Ars Rhetorica*. Oxford: Clarendon Press

- (1973): *Aristotelis Physica*. Oxford: Clarendon Press

- (1978): *Aristotelis Anaytica priora et posteriora*. Oxford: Clarendon Press

- (1979): *Aristotelis Topica et sophistici elenchi*. Oxford: Clarendon Press

Russel, P. (1995): *Freedom & Moral Sentiment: Hume's Way of Naturalizing Responsibility*. New York: Oxford University Press

Sauvé-Meyer, S. (1988). 'Why involuntary actions are painful'. *Southern Journal of Philosophy* (Supplement) 27, 127-158

- (1993): *Aristotle on Moral Responsibility: character and cause*. Oxford: Basil Blackwell.

- (1998): 'Moral responsibility: Aristotle and after'. In Everson, S. (Ed), *Companions to Ancient Thought, Volume 4: Ethics* (pp. 221-240). Cambridge: Cambridge University Press

Sharples, R. (1982): *Alexander of Aphrodisias on Fate*. London: Duckworth

Sher, G. (2006): *In Praise of Blame*. New York: Oxford University Press

Siegler, F. A. (1968): 'Voluntary and involuntary'. *Monist*, 52, 268-87

Smart, J. J. C. (1961): 'Free will, praise and blame'. In Watson, G. (Ed.): 2003, pp. 58-71. (Reprinted from *Mind* (pp. 291-306), 70)

Smith, H. (1983): 'Culpable ignorance'. *The Philosophical Review* 92, 543-571

Solomon, J. (1915): *The Eudemian Ethics*. In Ross, D. (Ed.), *The Works of Aristotle Translated into English, Volume 19*. Oxford: Oxford University Press.

Sorabji, R. (1973-4): 'Aristotle on the role of intellect in virtue'. In Rorty, A. O. (Ed.): 1980, pp. 201-220. (Reprinted from *Proceedings of the Aristotelian Society* 74, 107-129)

- (1980): *Necessity, Cause and Blame: perspectives on Aristotle's theory*. London: Duckworth

Speight, A. (2005): '"Listening to reason": The role of persuasion in Aristotle's account of praise, blame, and the voluntary'. *Philosophy and Rhetoric* 38, 213-225

Statman, D. (Ed.) (1993): *Moral Luck.* Albany: State University of New York Press

Stewart, J. A. (1892): *Notes on the Nicomachean Ethics of Aristotle, Volume 1.* Oxford: Clarendon Press

Stinton, T. C. W. (1975): 'Hamartia in Aristotle and greek tragedy'. *The Classical Quarterly* 25, 221-254

Stocker, M. (1986): 'Dirty hands and conflicts of values and of desires in Aristotle's ethics'. *Pacific Philosophical Quarterly* 67, 36-61

Strawson, P. F. (1962): 'Freedom and resentment'. In Watson, G. (Ed.): 2003, pp. 72-93. (Reprinted from *Proceedings of the British Academy* 48, 187-211)

Susemihl, F. (1883): *Eudemi Rhodii Ethica Aristotelis Ethica Eudemia*. Leipzig: Teubner

Taylor, C. C. W. (2006): *Aristotle: Nicomachean Ethics Books II, III and IV.* Oxford: Clarendon Press

Tuozzo, T. M. (1997): 'Aristotle as modern moral philosopher'. *Phronesis* 42, 335-344

Urmson, J. O. (1987): *Aristotle's Ethics.* Oxford: Basil Blackwell

Walker, M. U. (1993): 'Moral luck and the virtues of impure agency'. In Statman, D. (Ed.): 1993, pp. 235-250

Wallace, J. D. (1974): 'Excellence and merit'. *The Philosophical Review* 83, 182-199

Wallace, R. J. (1994): *Responsibility and the Moral Sentiments*. Cambridge (MA), London: Harvard University Press

Watson, G. (1975): 'Free agency'. In Watson, G. (Ed.): 2003, pp. 337-351. (Reprinted from *Journal of Philosophy* 72, 205-220)

- (1996): 'Two faces of responsibility'. *Philosophical Topics* 24, 227-248

- (Ed.) (2003): *Free Will* (2nd ed.). New York, NY: Oxford University Press

Wiggins, D. (1978/79): 'Weakness of will, commensurability, and the objects of deliberation and desire'. In Rorty, A. O. (Ed.): 1980, pp. 241-266. (Reprinted from *Proceedings of the Aristotelian Society* 79, 251-277)

Williams, B. (1990): 'Voluntary acts and responsible agents'. *Oxford Journal of Legal Studies* 10, 1-10.

- (1993a): 'Moral luck'. In Statman, D. (Ed.): 1993, pp. 35-56

- (1993b): *Shame and Necessity*. Berkeley, Los Angeles: University of California Press

- (1995): 'Internal reasons and the obscurity of blame'. In his *Making Sense of Humanity and Other Philosophical Papers* (pp. 35-45). Cambridge: Cambridge University Press

Williams, C. J. F. (1985): 'Aristotle's theory of descriptions'. *The Philosophical Review* 94, 63-80

Wolf, S. (1990): *Freedom within Reason*. New York, Oxford: Oxford University Press

Woods, M. (1992): *Aristotle: Eudemian Ethics Books I, II and III* (2nd ed.). Oxford: Clarendon Press

Zanuzzi, I. (2007): 'A definição do voluntario na Ethica Eudemia II.9'. *Journal of Ancient Philosophy* 1 (2), no pagination

Zingano, M. (2009): 'L'acte volontaire et la théorie aristotélicienne de l'action'. *Journal of Ancient Philosophy* 3 (2), no pagination