

ARISTOTLE'S
Ethics and
Moral Responsibility

JAVIER ECHEÑIQUE

CAMBRIDGE

CAMBRIDGE

more information - www.cambridge.org/9781107021587

ARISTOTLE'S *ETHICS* AND MORAL RESPONSIBILITY

Aristotle's *Ethics* develops a complex theory of the qualities which make for a good human being, and for several decades there has been intense discussion about whether Aristotle's theory of voluntariness, outlined in the *Ethics*, actually delineates what modern thinkers would recognise as a theory of moral responsibility. Javier Echeñique presents a novel account of Aristotle's discussion of voluntariness, arguing – against the interpretation by Arthur Adkins, and that inspired by Peter Strawson – that he developed an original and compelling theory of moral responsibility and that this theory has contributed in important ways to our understanding of coercion, ignorance and violence. His study will be valuable for a wide range of readers interested in Aristotle and in ancient ethics more broadly.

JAVIER ECHEÑIQUE is an associate professor at the Institute of Philosophy at the Pontifical Catholic University of Chile.

ARISTOTLE'S *ETHICS* AND MORAL RESPONSIBILITY

JAVIER ECHEÑIQUE



CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9781107021587

© Javier Echeñique 2012

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2012

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data

Echeñique, Javier, 1980–

Aristotle's Ethics and Moral Responsibility / Javier Echeñique.

pages cm

Includes bibliographical references and index.

ISBN 978-1-107-02158-7

1. Aristotle – Ethics. 2. Ethics, Ancient. I. Title.

B491.E7E24 2012

171'.3 – dc23 2012006478

ISBN 978-1-107-02158-7 Hardback

Cambridge University Press has no responsibility for the persistence or
accuracy of URLs for external or third-party internet websites referred to
in this publication, and does not guarantee that any content on such
websites is, or will remain, accurate or appropriate.

Contents

<i>Acknowledgements</i>	page vi
<i>Abbreviations of Aristotle's texts</i>	viii
Introduction	I
PART I TOWARDS AN ACCOUNT OF ETHICAL ASCRIPTION	
1 Against the Strawsonian interpretation	19
2 Aristotle on ethical ascription	55
PART II THE DEFEATERS OF ETHICAL ASCRIPTION	
3 The definitions of violence	87
4 Coercion as justification and excuse I: the <i>Ethica Eudemia</i>	112
5 Coercion as justification and excuse II: the <i>Ethica Nicomachea</i>	134
6 Factual error and the source of blame	148
7 The pain condition	173
Conclusion	192
Appendix: Two lexical ambiguities	195
<i>Bibliography</i>	199
<i>Index</i>	207

Acknowledgements

This book is the result of a conglomerate of efforts, not entirely my own. First and foremost, it is the result of working under the tutelage of Sarah Broadie, to whom I am particularly indebted for her sustained support during my years of study at the University of St Andrews, for disciplining what I initially thought of as incorrigible argumentation, and for provoking many ideas. It is also the result of having reflected on the insightful criticisms and suggestions of David Charles and Anthony Duff, as well as on the comments of those who have read individual chapters (Jens Timmerman, Dan Labriolla, Beris Gaut and James Harris).

I have also greatly benefited from having worked in a propitious academic environment during those years, one that gave me the opportunity of exposing my ideas in conferences (the APS 2010 and the NAAP 2008 conferences) and seminars, as well as in more informal contexts. My special thanks in this regard go to all the Chilean taxpayers who indirectly sustained my biological life throughout my doctorate, through the Presidente de la República Scholarship.

After I submitted my Ph.D dissertation, a number of people also read and commented on drafts of this book, or parts of it. Timothy Chappell, who read and commented on the book for Cambridge University Press, made many detailed suggestions for improvement, and two anonymous referees for Cambridge University Press also contributed to sharpening my perception of the possible objections that could be made to some of my arguments. I should like to record my gratitude to all the people mentioned above.

The work of turning the dissertation into a book has been mostly carried out in Santiago de Chile, as a lecturer at the Centre of Classical Studies (Universidad Metropolitana de las Ciencias de la Educación) and at the Department of Philosophy, Universidad de Chile. I am grateful to all my colleagues for their kind support and for the professional environment they

help to create, despite the challenging circumstances surrounding public education in Chile. This book is dedicated to all of them.

To Hilary Gaskin at Cambridge University Press I am grateful for believing in this project from the very beginning, and to Anna Lowe, Tom O'Reilly and Elizabeth Davison for their diligence and kindness in answering all my queries.

I am also obliged to Angela Viola-Glapinska for tolerating during this year the inevitable side-effects associated with the writing of a book, and to my dear parents, Jorge and Rosario, for offering me their firm support during the whole period of my interest in philosophy, since its uncertain beginnings when I was an undergraduate student.

As will become evident, this work is also greatly indebted to Susan Sauvé Meyer's book *Aristotle on Moral Responsibility*, which has considerably raised the standards of discussion on this topic. *Aristotle's Ethics and Moral Responsibility* was completed before the reissue of Suavé Meyer's book in 2011, so references are to the 1993 edition.

Abbreviations of Aristotle's texts

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 Motu Animalium: *MA.*
De Sophisticis Elenchis: *Soph. El.*
De Virtutibus et Vitiis: *De Virt.* (spurious)
Ethica Eudemia: *EE*
Ethica Nicomachea: *EN*
Magna Moralia: *MM* (disputed authorship)
Metaphysica: *Met.*
Physica: *Phys.*
Poetica: *Poet.*
Politica: *Pol.*
Topica: *Top.*

Introduction

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?

G. E. M. Anscombe, ‘Modern moral philosophy’, 381

A. REPUTABLE OPINIONS

Aristotle’s two ethical treatises, the *Ethica Nicomachea* (*EN*) and the *Ethica Eudemia* (*EE*), centre around the practical question ‘What is the primary good for human beings as such?’, and they can be succinctly characterised as systematic efforts to work out an answer to this question, and to consider its implications. Aristotle’s proposal in his *Ethics*¹ embraces both the common opinion that the chief human good is *eudaimonia*, wellbeing or flourishing, and his own peculiar conception of this purely laudatory notion, as activity of the soul in accordance with excellence in a complete life (*EN* 1098a7–18; *EE* 1219a34–9). This definition sets the agenda of Aristotle’s *Ethics*, which develops into a complex theory of the human *aretai* or excellences, that is, those qualities that make for a good human being. This book is about an integral aspect of this theory: Aristotle’s discussion of voluntariness (*to hekousion*), which in both ethical treatises, the *Ethica Eudemia* and the *Ethica Nicomachea*, precedes the discussion of the various human excellences (*EN* III 1; *EE* II 6–9).

¹ From now on I will use the term ‘Ethics’ in italics to designate Aristotle’s two ethical treatises, the *Ethica Nicomachea* (*EN*) and the *Ethica Eudemia* (*EE*). I will occasionally make reference to the *Magna Moralia* (*MM*), but I think none of my interpretations depend on attributing this work to Aristotle himself. The reader should also bear in mind that Books IV–VI of the *Ethica Eudemia* also appear as Books V–VII of the *Ethica Nicomachea*. Again, none of my interpretations depend on identifying the *Ethica Eudemia* as the first home of the common books (although I am inclined to do so).

Without any doubt, the most debated question with regard to Aristotle's discussion of voluntariness in his *Ethics* has been whether it is concerned with delineating a theory of *moral responsibility*, perhaps the first theory of its kind in the history of Western philosophy. In his 1960 book, *Merit and Responsibility*, in all probability the first systematic effort in the English language to closely scrutinise the development of this notion in ancient Greece, Arthur W. H. Adkins famously came to a negative conclusion in answer to this question: Aristotle fails to develop a satisfactory concept of moral responsibility, for much 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',² was the traditional conception of *aretê* (excellence or virtue). According to Adkins, this conception 'was inimical to the concept of moral responsibility'.³ The second premise is simply the application of this general thesis to Aristotle: Aristotle inherits this traditional conception of *aretê*. This conception embraces, in addition to what Adkins terms the 'quiet', 'cooperative' 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 conception of *aretê* was inimical to the concept of moral responsibility? If I interpret him correctly, Adkins believed that the traditional conception 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.⁴

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

² Adkins 1960: 332. ³ *Ibid.* 337. ⁴ *Ibid.* 2.

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 related 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*, *authority*, etc., relate to one another. 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 characterised 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, whenever an agent carries out a morally significant action for which he is responsible, he thereby renders himself amenable to praise and blame, that is, to moral assessment or evaluation.⁵ This much is uncontroversial. Nonetheless, controversy begins to arise as soon as we attempt to elucidate the nature of this moral assessment, for it is clear that a given conception of moral responsibility will partly depend on what we take the nature of this assessment or evaluation to be, and the latter will in turn determine what pieces of behaviour are relevant to the assessment of an agent, and more importantly, *what sort of agents* are fitting objects of moral assessment. The insight peculiar to the post-Strawsonian debate in this regard is (2) the increasing recognition that, what I shall 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 P. F. Strawson, some philosophers like C. L. Stevenson and 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* – and this mere socially useful grading makes sense only when the actions being graded are voluntary. After P. F. Strawson's 'Freedom and resentment', philosophers have widely come to the conclusion 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

⁵ When referring to 'the agent' in my examples, or 'S', I will be using the masculine pronoun.

this agreement is that this purely socially useful grading ‘instrumentalises’ the sane, adult human agent and fails to acknowledge him 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 *x* (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 doing *x*. Essential to this conception of moral responsibility is 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](#)), as well as 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 pure socially useful grading fails to do justice to the moral nature of our moral praising and blaming practices, is 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 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 affected by 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’. As a result of this fusion, the Strawsonian wrongly categorises *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 [Chapters 1](#) and [2](#) depends on showing that the two most influential post-Strawsonian interpretations of Aristotle’s theory of voluntariness inherit

the flaws of the Strawsonian account; in particular, the error of dismissing educational attitudes as non-moral, and thus pupils in the process of moral education as non-moral agents. As is natural, these two (mutually excluding) post-Strawsonian views are ultimately based on an interpretation of the sort of praise and blame which, in Aristotle's view, is bestowed only upon voluntary responses (*EE* 1223a9–15; *EN* 1109b30–4).

One such interpretation is what I call the 'prospective conditioning interpretation' about the significance of the involuntary/voluntary distinction for Aristotle's *Ethics*. According to this interpretation, the sort of praise and blame Aristotle says is appropriate to voluntary responses is *prospective* in nature.⁶ 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-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) the praise and blame in (ii) are not moral attitudes, and their recipients are not moral agents. This, I think, is the argument underlying Jean Roberts' 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'.⁷ 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*: the sort of praise and blame that an Aristotelian agent 'deserves' for those responses he is *causally responsible* for, is based on a retrospective evaluation, 'for the question is *whether the agent's character produced them*'.⁸

⁶ This view is chiefly represented by Roberts 1989. ⁷ *Ibid.* 25.

⁸ Sauvé Meyer 1993: 51 (emphasis added).

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/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 retrospective evaluations of mature ethical agents in possession of ethical dispositions. By ‘retrospective evaluations’ Sauvé Meyer means that these are evaluations of ‘morally’ significant pieces of behaviour that can be traced back to the corresponding ethical dispositions 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? How is this emphasis on the causal antecedent supposed to help us understand why our evaluating an agent in possession of a character trait for having caused a given response is a distinctively moral evaluation, in a substantive sense of ‘moral’? It is clear that Sauvé Meyer and Irwin rely 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.

The maturing of these ideas furnished the starting point for this book, and they are developed in some detail in [Chapter 1](#). In [Chapter 2](#), I develop my own positive interpretation of what I take to be Aristotle’s concern with the voluntary/involuntary distinction in his *Ethics*. Another insight of the post-Strawsonian debate (this time, Gary Watson’s), turned out to be extremely significant in this regard: (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 evaluations bestowed on human agents on the occasion of their conduct that are made from a different perspective and do 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. Watson calls this sort of appraisal ‘attributability’, that is, the mere attribution of an action to an agent.

If Aristotle’s concerns are indeed better modelled on a conception of responsibility as attributability, as I think they are, it follows that 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’. 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. In their focal use, praise/blame *logoi* apply to the prohairetic agent, the fully virtuous or faulty agent capable of *prohairesis*, rational choice. When I call someone S ‘just’ or ‘cowardly’ on the occasion of his recognisably 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 captures the truth contained in the special emphasis attached to the prohairetic agent in the Strawsonian account. Moreover, this account of praising and blaming *logoi*, as having both focal and 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.

The domain of moral agency appropriate to ethical ascription, I further argue in [Chapter 2](#), is best characterised by the notion of reason-responsiveness. The reason-responsive agent is 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 (i.e. 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 (i.e. the pain-responsive agent, a category that includes non-rational animals and irreversible psychopaths).

B. CONDITIONS OF ETHICAL ASCRIPTION AND THE NEGATIVE METHOD

This book has two parts, and the preceding section pretty much summarises the contents of the first one, 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 x as ‘V’ and [2] praising or blaming a reason-responsive agent S in ‘V’-terms through (*dia*) [1]?

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 call ‘moral pain’ (i.e. [2]). The difference between these two conditions will be the subject matter of Chapter 7.

The conditions of praise and blame are not to be confused with the narrower set of conditions that, according to Aristotle, an action x (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, x needs not only (i) to be voluntarily performed (e.g. done by yourself in full knowledge that what you are refusing is to eat an enormous cake, a recognisably temperate action) and (ii) done without the relevant sort of pain, and perhaps with pleasure (EN 1104b5–8); x must also (iii) be the result of *prohairesis*, a considered judgement or decision to the effect that x is, overall, good/bad conduct (EN 1105a31), and (iv) good/bad conduct *qua* x (EN 1105a32), e.g. refusing to eat it *qua* recognisably 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.

⁹ ‘Full’ or ‘perfect’ ethical dispositions are not meant to include continence and incontinence.

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* III b9), 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: ‘From 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. According to Furley, in his *Ethics* Aristotle ‘has to find the distinguishing characteristics of a subset of animal movements – namely, human actions *for which we hold the agent morally responsible*’.¹² Indeed, as we shall see in [Chapter 2](#), 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*, the voluntary acquisition of such a disposition. Indeed, no

¹⁰ They act voluntarily (*EN* II 5 215–19), however. Aristotle thinks that the incontinent is not fully faulty because his rational choice (*prohairesis*) is good; it is just that he fails to act on it.

¹¹ Furley 1967: 191. See also Furley 1978: 62–5. More recently, the idea that responsibility for one’s actions requires responsibility for one’s character has been assumed (without argument) by Pakaluk 2005: 122 (where he says that, unless we assume responsibility for character, ‘we cannot hold a coward responsible for his cowardly action’, emphasis added), and also by Destrée 2011.

¹² Furley 1978: 62 (emphasis added).

such argument for the voluntary acquisition of ethical dispositions is to be found in the *Ethica Eudemia*.

About condition (ii), the lack of pain condition and its contrary, the presence of pain, let me just say that these conditions involve rather sophisticated, focused attitudes towards the ethically salient features of one's practical situation, and that, although this sort of focused lack of pain and its contrary certainly require a certain level of moral development, there is no reason to think that 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 [Chapters 3](#) and 7.

What about condition (i) concerning voluntariness? There are strong reasons to doubt 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 *involuntariness*. This negative approach is present in the *Ethica Eudemia* and the *Ars Rhetorica*, and is conspicuous in the *Ethica Nicomachea*. The *Ethica Eudemia*, 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). This negative notion of the voluntary is reflected in the Eudemian definition of the voluntary as (a) what the agent does 'not in ignorance' (*mê agnoôn*, 1225b8–9). This same definition includes a second condition, namely, that what the agent does 'not in ignorance', he also does (b) 'through his own agency' (*di' auton*, 1225b9). It is clear, however, that Aristotle has arrived at this notion of acting through one's own agency via the previous analysis of violence (*bia*) in *EE* II 8: 'acting through violence *and not through violence* seem to be akin to the notions mentioned [the involuntary and the voluntary]' (1224a10–11). The *Ars Rhetorica* makes this latter point explicit, by defining voluntary actions as 'those things that we do with knowledge and *without violence* (*mê anagkazomenoi*)' (1368b9–10). *EN* III 1 sums up the negative strategy thus: '*given that* the involuntary is <what is done> through violence or factual unawareness, one could think that voluntary is that action the source of which is in one self, while one is aware of the particular circumstances in which it occurs' (*EN* IIIa22–4). The conditions of voluntariness, it seems, can be regarded as the negations of the conditions of involuntariness.

Aristotle goes further than this, however, for he clearly assumes that an analysis of the conditions of involuntariness, such as violence and factual error, is a good method for arriving at genuinely *positive* conditions of voluntariness, like the condition of internal origin and factual knowledge. As I show in [Chapter 3](#) (sections D to E), there are good reasons to doubt the value of this assumption. The more restricted negative approach that I follow, on the contrary, 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 characteristics of actions, but rather concepts that, when applied to actions, ‘rule out the suggestion of some or all of its recognized antitheses’. 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 briefly explain this.

I am well aware that the notion of the voluntary plays a role in *De Motu Animalium* (*De Motu* 703b3), where it is used in the physio-psychological context of a mechanical explanation of animal motion. Moreover, Aristotle explicitly says in *EN* 1111b8–9 that non-human animals share in voluntary action (cf. 1111a24–6). Aristotle, however, clearly thinks that non-human animals are not appropriate targets of praise and blame, because their behaviour cannot be appropriately assessed in virtue or vice terms (*EN* 1145a25–7, 1149b31–5). Now, if ‘*hekousion*’ (‘voluntary’) is sometimes used without reference to praise and blame, one possible view is to take the *basic* extension of the term (when applied to actions) to be somehow identifiable independently of our practices of praising and blaming. The *De Motu* tacit characterisation of voluntary animal motions as those motions (in Nussbaum’s words) ‘to explain which we have to mention the agent’s desire and his view (his *phantasia*) of the goal’, is the obvious candidate.¹⁴ On this view, Aristotle in his *Ethics* would be concerned only with that sub-category of voluntary actions thus characterised, namely, those that are open to praise and blame.¹⁵

There are at least two considerations that count against this suggestion, however. First, there is the fact that ‘voluntary’ (*hekousios*) as predicated of animal motion, and without reference to praise and blame, only occurs *twice* in the Aristotelian *corpus* (once in *De Motu Animalium*, and once

¹³ Austin 1956–7: 8.

¹⁴ Nussbaum 1978: 380. See *De Motu* 703b2ff. Here Aristotle is concerned with distinguishing certain sorts of animal motions from the sudden stimulus-occasioned responses of particular parts of the body, like the heart, and natural bodily processes like respiration.

¹⁵ I would like to thank David Charles for this suggestion.

in the *Ethica Nicomachea*), by contrast with the myriad instances where this notion, in its ethical use (i.e. as ‘openness to assessment’) is found in the *Ethica Eudemia*, the *Ethica Nicomachea*, and the *Ars Rhetorica*. If the basic extension of ‘*hekousion*’ can be fixed independently of our practices of praise and blame, then why isn’t this basic sense more prominent in the Aristotelian *corpus*? Secondly, there is also the fact that the definitions of voluntariness to be found in these texts (indeed, the *only* explicit definitions of voluntariness to be found in the *corpus*) are arrived at via an examination of standard *excuses* like factual ignorance and violence. This is important because there simply is no other criterion to regard conditions like factual error and violence as conditions that render an action *involuntary*, apart from the fact that they render the action an inappropriate target of praise and blame.¹⁶

These two considerations strongly suggest that the notion of voluntariness, as predicated of an agent *S* (*hekôn*) or an action *x* (*hekousia*), is primarily used to mark the fact that *S* or *x* is open to evaluation, i.e. open to praise or blame. In general, I am quite sympathetic 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’.¹⁷ It seems to me that the voluntary (*to hekousion*) is a case at hand, and accordingly that the use of this notion to characterise certain sorts of animal motion, far from being basic, is derivative.

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 *closed* to assessment is present, is tantamount to characterising them in terms of

¹⁶ Someone may object to this assertion by pointing out that Aristotle in *EE* II 8 attempts to ground the concept of violence (*bia*) on the wider concept of contrariety to nature: in this sense, the upward motion of a stone can be said to be ‘forced’ because the stone’s natural impetus is to travel downwards. Stones do not engage in ethically significant behaviour, however. But Aristotle recognises that ‘violence’ (*bia*) is predicated of elemental motion only ‘because of a certain resemblance’ (*EE* 1224b4–5). What then is the criterion for applying the term ‘violence’ *sensu stricto*? It seems that the resemblance in question cannot just be resemblance to a ‘basic’ sense of violence as ‘what is against an organism’s natural impulse’ (the counterpart of the supposedly ‘basic’ sense of the voluntary), for this would be arbitrary: why stop at an *organism*’s natural impulse? Why not just say that any motion that is contrary to a *natural entity*’s natural impetus is violent, full stop? The resemblance in question, as the context of *EE* 1224b4–5 shows, must be the resemblance that contrariety to an inanimate impetus bears to a context where violence functions as an excuse.

¹⁷ Williams 1993b: 42.

the voluntary. To this, one must also add the fact that the practice of attributing an ethically significant action x to a reason-responsive agent is usually not consequent upon first establishing that x is voluntary because none of the conditions that render it closed to assessment are present. Such practice rather takes 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,¹⁸ 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.

C. DEFEATERS OF ETHICAL ASCRIPTION

In the second part of this book we shall be concentrating our efforts on the factors that defeat the ascription of ethically significant behaviour to reason-responsive agents: the conditions of involuntariness and the pain condition. 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 attend 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 such defeaters, 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 by Aristotle: (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 can be an important basis for an ethical assessment of S , in so far as these attitudes express S 's evaluative outlook. In Chapter 7 I argue that Aristotle recognises the significance of such attitudes. In particular, for Aristotle the presence of a moral-aversion attitude figures as defeater 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

¹⁸ 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.

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 *internal compulsion*. So, apparently Aristotle's two simple and seemingly 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 exempting conditions, that is, conditions like psychopathy that render *an agent* an unsuitable recipient of praise and blame in a sustained manner. Admittedly, Aristotle never attempts to answer the question: 'What is it about an organism that makes it an appropriate recipient of praise and blame?' Nevertheless, Aristotle does recognise exempting conditions, and characterises agents in abnormal psychological states.¹⁹ Furthermore, I suggest in [Chapter 2](#) that Aristotle has a principled answer to the aforesaid question, in terms of reason-responsiveness.

(C) Now, conditions such as violence, internal compulsion and factual error seem to amount to mere excuses, by showing that a bad action was performed involuntarily. But certainly one should also expect Aristotle to recognise the particular role of *justification* – only now not as a kind of ascription-defeater, but rather as a blame-defeater – in the sense of reasons to believe that, although one did the *prima facie* wrong action voluntarily, one was *justified* in doing it, so that one is not an appropriate target of blame. 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 them, if it was the agent's *own fault*

¹⁹ See, for instance, the *EN* VII 5 discussion of exempting conditions like disease or madness.

that he put himself in a situation where any of them 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 to F) that this objection is based on a false assumption.

If these four ways of ‘upgrading’ the conditions defeating ethical ascription are acknowledged, the resulting account turns out to be surprisingly rich and appealing. Aristotle’s account is also internally consistent, and even though there are differences between the Eudemian and the Nicomachean accounts, such differences are either apparent, or they can be understood as part of a coherent logical development where the Nicomachean account improves on the Eudemian. If by the end of [Chapter 7](#) the reader has come to appreciate these arguments, I believe the purpose of the second part of this book has been fulfilled.

PART I

Towards an account of ethical ascription

Against the Strawsonian interpretation
The negative argument

A. VOLUNTARINESS, PRAISE AND BLAME

In *Ethica Eudemia* II 6, where Aristotle introduces the voluntary/involuntary distinction (*hekousion/akousion*), he employs four expressions (with the same referent 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 (*EE* 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 (1223a5, a7); (iii) these actions of which he is a master and the source depend on him (*eph' hautô[i]*) (1223a7); finally, he claims – in what seems to be a terminological digression aimed at defining these concepts in terms of each other – that (iv) the human being is *aitios* for/of these same actions (1223a8, 15, 16). The latter concept is obviously related 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 Apparently, Aristotle believes that any of these notions, including of course the notion of voluntariness (*to hekousion*), could be substituted for any other *salva veritate* (*EE* 1223a4–7, 1223a7–9, 1223a17–20; *EN* IIIb20–I, 1136b27–8).

Now, Aristotle explicitly says that a source (*archê*) of the becoming of *x* is always a cause (*aitia*) of *x*, and indeed the *first cause* (*EE* 1222b31). Hence, when he says that ‘excellence and vice have to do with those actions of which one is oneself *aitios* and the source’ (*EE* 1223a15–16), ‘*aitios*’ (the adjective) seems to convey more than the idea of *cause* (*aitia*, the 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

¹ *EN* 110a1, 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]*.

as ‘cause of x ’, as translators tend to do.² 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 accused in legal contexts, and *hê aitia* the accusation). This temptation notwithstanding, the major problem with translating ‘*aitios*’ as ‘responsible’ in these contexts is that the English term bears strong nuances of *accountability* or *answerability*³ (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 the aforementioned causal notions reflects a concern with accountability or answerability. It is therefore important to avoid prejudging the nature of Aristotle’s concern with these notions in these contexts, and so I will try to avoid talk of ‘responsibility’ or ‘moral responsibility’ and cognate notions (indeed, I will leave ‘*aitios*’ untranslated), unless I am reporting particular views that make explicit use of them.

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, with praise (*epainos*) and blame (*psogos*). According to the *Ethica Eudemia*:

Since virtue and vice and the operations that spring from them are respectively praiseable and blameable (for what is due to necessity, chance, or nature, is not praised or blamed, but only 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 with (*peri*) those actions of which he is himself *aitios* and a source. We must therefore grasp the sort of actions of which he is himself *aitios* and a source. (*EE* 1223a9–16)⁴

These actions, the *Ethica Eudemia* goes on to say (a17), are voluntary (*hekousia*), so we must study them. The opening lines of *EN* III state a very similar reason for pausing over the concept 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* 1109b30–4)

² Examples are H. Rackham, M. Woods and J. Solomon.

³ In fact, the Latin term ‘*responsus*’ (from which the English ‘responsible’ ultimately derives) is the participle of ‘*respondere*’, ‘to answer’.

⁴ Unless otherwise indicated, quotations from Aristotle are translated by me.

According to the *Ethica Eudemia* and the *Ethica Nicomachea*, the reason why voluntariness is a topic of significance for the *Ethics* can be thus summarised: given that Ethics is an enquiry about the highest good for human beings as 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* (and derivatively, of faulty activity). This is the implicit background of the passages. In both treatises⁵ Aristotle's reasoning seems to be that, since (1) virtue and vice and the responses virtue and vice are *concerned with* (*peri*) are praiseable and blameable,⁶ respectively (assumed in *EN* 1109b30–4), and (2) praise and blame are bestowed upon voluntary responses, responses of which we are *aitioi*, then (3) the responses virtue and vice are concerned with (i.e. ethical activity) must be *voluntary* responses, and so we must study voluntariness.

Now, the phrase 'affections and actions that are the *concern* of virtue and vice' in premise (1) is ambiguous between (A) the kind of responses by which excellences and faults of character *come about*, which involve voluntary actions (*EN* 1113b26–7, 1114a12–16); and (B) the kind of responses that ethical dispositions *dispose us to do*. This ambiguity of 'ethical activity' between (A) and (B) is promoted by Aristotle's belief that ethically significant responses 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, in accordance with themselves (*kat' autas*)' (*EN* 1114b26–7).

The crucial claim, on the basis of which many interpreters have tried to resolve the aforementioned ambiguity, is Aristotle's suggestion that the way to arrive at the conditions of virtuous (and faulty) activity is by specifying the conditions of our *practices of praising and blaming*, and in particular, the fact that these practices are targeted at ethically significant activities that are voluntary, activities of which we are *aitioi*. On the basis of certain modern accounts of the nature of praise and blame, some interpreters (the 'Strawsonian interpreters') have argued that Aristotle's discussions of voluntariness in his *Ethics* are exclusively concerned with the conditions of virtuous (and faulty) activity of type (B), that is, the type of voluntary activity that virtue (or vice) disposes us *mature ethical agents* to do. These interpreters have

⁵ See also *MM* 1187a20–1.

⁶ For reasons that will become clear in section F below, I prefer to translate '*epainetos*' and '*psektos*' as 'praiseable' and 'blameable', respectively, instead of the more standard 'praiseworthy' and 'blameworthy'.

also argued 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. By the end of [Chapter 2](#), I expect to have shown that Aristotle is also concerned with type (A) responses.

B. THE STRAWSON-INSPIRED INTERPRETATION OF ARISTOTLE

What emerged from the previous section is that an appropriate understanding of Aristotle's notion of praise and blame is of paramount importance for understanding the nature of Aristotle's concerns with voluntariness in his *Ethics*. Now, our praising and blaming practices are also important to the most influential account of moral responsibility in contemporary Anglo-American philosophy, namely, the one developed by P. H. Strawson in his famous paper 'Freedom and resentment'.⁷ In this paper, Strawson focuses on the attitudes and feelings that human beings manifest to each other whenever they believe that holding one another responsible is appropriate. Strawson calls these 'reactive attitudes', and they include attitudes and feelings such as resentment, indignation, thankfulness, approbation, guilt and pride. These attitudes have always been recognised in one way or another in connection with moral responsibility, but they have traditionally been seen as mere corollaries of some independently intelligible concept of moral responsibility. The peculiarity of Strawson's view is that the reactive attitudes are *constitutive* of moral responsibility.

According to Strawson, for S to be morally responsible for an item of behaviour *x* (or the consequences of *x*) just *is* for S to be the *appropriate target* of reactive attitudes on the part of what Strawson calls 'the members of the moral community', on the occasion of his doing *x*: in particular, it is for S to be the appropriate target of resentment on the part of the affected parties, or moral indignation on the part of someone who is not directly affected, if *x* is an evil; alternatively, it is for S to be the appropriate target of gratitude on the part of the beneficiary, or moral respect on the part of someone who is not directly benefited by *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

⁷ Strawson 1962. For a careful development of the Strawsonian model of moral responsibility, see Wallace 1994. For a discussion of the historical roots of the Strawsonian model in Hume's philosophy, see Russell 1995.

party, as resentment and gratitude are; or impersonal, 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 the following conditions hold: (i) first, S must be a member of the moral community, that is, S must be himself capable of holding the other-reactive attitudes and also *self-reactive* ones like guilt, compunction or remorse,⁸ along with other capacities necessary for holding these (e.g. the capacity to put oneself in others' shoes). According to Strawson, children and psychopaths are denied membership in the moral community precisely because they lack these capacities.⁹ (ii) Secondly, x must be such that it breaches (if the attitudes are negative) or complies with (if they are positive) what Strawson describes as 'an expectation of, and demand for, the manifestation of a certain degree of goodwill or regard on the part of other human beings',¹⁰ towards ourselves and other human beings. Strawson thought that this goodwill that we human beings demand from one another is an essential element of our interpersonal relationships and our moral life as we know it, and it is what ultimately grounds the reactive attitudes. 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 or 'moral obligation'.¹¹ (iii) Finally, x must be voluntary. This condition captures the fact that someone's goodwill or regard for others is only expressed in that behaviour of his which is voluntary. As a corollary of this, the reactive attitudes towards S for his having done x are inhibited whenever x is shown to be involuntary.

This is a very sketchy account of the Strawsonian conception, and some important gaps will be filled in later in this section and in section F below. Nevertheless, this rough sketch should suffice to make the following point. Because Aristotle maintains 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) and because of the post-Strawsonian temptation to regard Aristotelian

⁸ These are what 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 So Strawson says that an agent is excluded from the moral community, that is, from the domain of other-regarding reactive attitudes, if he 'is seen as excluded from ordinary adult human relationships by deep-rooted psychological abnormality – or simply by being a child' (Strawson 1962: 81).

¹⁰ *Ibid.* 84. ¹¹ *Ibid.* 86.

praise and blame as S-reactive attitudes¹² and to find in Aristotle's *Ethics* criteria that appropriately distinguish moral from non-moral agents in the Strawsonian sense, it is possible to think that, according to Aristotle, for S to voluntarily carry out a blameable or praisable action *x* and for him to be blameable or praisable 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 blameable or praisable for having voluntarily done *x* is, I shall argue, on the wrong track. But first we must take a closer look at the main features of the Strawsonian interpretation.

[I] *Aristotelian praise and blame as S-reactive attitudes*

Terence Irwin, for instance, in attributing to Aristotle a 'theory of responsibility', makes it clear that by this expression he means 'an account of our practice of *holding 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.'¹³ Irwin explicitly points out that for him, 'holding people responsible' corresponds to this very notion of responsibility *à la* Strawson:

Aristotle says that *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 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.¹⁴

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. Furthermore, his including praise and blame in the list shows that he is interpreting Aristotelian praise and blame as S-reactive attitudes. More significantly, he clearly takes the S-reactive attitudes to be constitutive of the concept of being responsible.

The identification of praise and blame with the S-reactive attitudes is also evident in Susan Sauvé Meyer's book *Aristotle on Moral Responsibility*. Sauvé Meyer believes that Aristotle's account of voluntariness at *EE* II 6–9

¹² 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.

¹³ Irwin 1980: 133–4. ¹⁴ *Ibid.* 134 (emphasis added).

and *EN* III 3 ‘is intended to capture the conditions of moral responsibility for action’,¹⁵ 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.¹⁶

It is manifest that Sauvé Meyer also takes praise and blame to be two of the wide range of interpersonal attitudes that Strawson recognises as constitutive of the concept of responsibility. Moreover, it is clear that she agrees with Strawson precisely on their being constitutive of the concept of responsibility, for she thinks that moral responsibility is ‘the property of an agent that makes her subject to the demands of morality’, and equates ‘being subject to the demands of morality’ with ‘being subject to the S-reactive attitudes’. This is clearly the Strawsonian approach, not only in that the S-reactive attitudes are constitutive of moral responsibility, but also in that they are expressive of certain basic ‘moral demands’.

[III] Aristotelian recipients of praise and blame as accountable agents

A second feature of the Strawsonian interpretation follows immediately from the first, and can also be found hinted at in the two passages from Irwin and Sauvé Meyer just quoted. If the interpretation advanced by these scholars is in fact inspired by Strawson’s conception of moral responsibility, then it must provide the Aristotelian equivalent of Strawson’s condition (i) *supra*, namely, a characterisation, in Aristotelian terms, of the appropriate

¹⁵ Sauvé Meyer 1993: 6.

¹⁶ *Ibid.* 18 (emphasis added). See also Sauvé Meyer 1988: 144: ‘*suggnômê* also requires the remission of other reactive attitudes besides blame – attitudes such as anger, ill will, and resentment’.

target of S-reactive attitudes. It is unsurprising, therefore, that Irwin characterises (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).¹⁷ 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* [an eminently Strawsonian phrase], and those who are not'.¹⁸ 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, characterised by him as a *hexis prohairetikê*, a capacity of rational choice (*EN* 1106b36); indeed, one of Aristotle's requirements for virtue and vice is the *capacity of reason*, and in particular, *prohairesis*, rational choice.

The Strawsonian interpreter is thus able to cope with Aristotle's claim that animals and children are also capable of voluntary action (*EN* 1111a25–6). Since neither animals nor children are, as Strawson observes, the 'appropriate recipients' of S-reactive attitudes, nor do they have the competences to hold genuine S-reactive attitudes towards others, and since neither animals nor children are, according to Aristotle, capable of rational choice (*EN* 1105a31),¹⁹ the Strawsonian interpreter provides us with Aristotelian

¹⁷ Irwin 1980: 134 (emphasis added). See also Irwin 1988: 341. Here he focuses on the mature rational agent whose decision 'will reflect the results of his deliberation about his happiness, and therefore will express the sort of attitude that a non-rational agent cannot take to himself' – notice that this conception of rational agency excludes children.

¹⁸ Sauvé Meyer 1993: 19 (emphasis added).

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) '[a]nimals and children act voluntarily', results in a contradiction (see Irwin 1980: 125). I think that Irwin makes a double mistake here. One is to include animals in Aristotle's 'simple' 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 children are not responsible for their actions. Irwin wants to say: 'Aristotle clearly assumes that . . . children are not subject to legal and moral sanctions' (*ibid.* 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

criteria underlying their exclusion from the category of (Strawsonian) moral agents.²⁰ Animals, psychopaths and children are all excluded from this category *for the same reason*.²¹

Admittedly, Strawson himself does not characterise the ‘morally responsible agent’ as someone who has the capacity of *prohairesis* or rational choice, or as having the capacity of reason. He certainly prefers to characterise the moral agent as a reasonable candidate for the S-reactive attitudes. Nonetheless, both Irwin and Sauvé Meyer clearly intend these two characterisations of the moral agent to be co-extensive. This should already give us pause: perhaps they are co-extensive, but why should we assume that Aristotle is interested in delineating the concept of the agent in possession of an ethical disposition partly *because* he is interested in the concept of the answerable agent? As we shall see, the fact that these interpreters take the Strawsonian characterisation of the moral agent in terms of the S-reactive attitudes as a starting point, explains why their enquiries are strongly guided by the question of whether and how Aristotle distinguishes the ‘morally responsible agent’ from a pre-established category of non-responsible agents that includes not only beasts and psychopaths, but also children.²²

[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 accordingly vital for the arguments I will offer against it. Strawson argued that paradigmatic agent-based, exempting conditions (as distinct from *excusing* conditions, which apply to particular actions) like childhood or psychopathy, inhibit the S-reactive attitudes in a global way by encouraging us to view the child

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 viewpoint of accountability, but not, as we shall see, from the point of view of attributability.

²⁰ The most significant difference between Terence Irwin and Susan 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; Sauvé Meyer 1993: 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 2007: 180, he confines praise and blame to ‘agents capable of election’).

²¹ Sauvé Meyer 1993: 23.

²² There are other interpreters who do not explicitly share the Strawsonian outlook but are nevertheless inclined to add this further requirement of rational choice (or *possession* of a moral disposition) to Aristotle’s theory of voluntariness, as a requirement that would render such theory a theory of *moral* responsibility. Some examples are Bobzien *forthcoming*; Sakezles 2007: 227; and Pakaluk 2005: 121–49. Some of the arguments I present in this chapter (and particularly the ones in section D below) also count against these views.

or the psychopath as *the sort of agent* who is not to be held accountable for complying with the demands of morality. These exempting conditions ‘make us see the agent as other than a morally responsible agent’.²³ But Strawson also thought (and this is better regarded as a different thesis of his) that these exempting conditions like childhood or psychopathy, promote instead an altogether different sort of attitude that he calls the ‘objective attitude’, a sort of attitude that Strawson regards as profoundly *opposed* to the 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 taken account, perhaps precautionary account, of; to be managed or handled or cured or trained; perhaps simply to be avoided.²⁴

Since this is a distinctively Strawsonian thesis, it seems 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 mere ‘objective attitudes’. It is therefore unsurprising to find Irwin arguing precisely this:

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 thinks it better not to act on them.²⁵

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

²³ Strawson 1962: 85. ²⁴ *Ibid.* 79.

²⁵ Irwin 1980: 134 (emphasis added). See also Irwin 1988: 340, where ‘objective’ praise and blame is defined as the ‘mere communication of pleasure and pain’ in people’s actions. See also Irwin 2007: 178.

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.²⁶

Notice that the Strawsonian contrast between ‘reactive’ and ‘objective’ attitudes is now framed, by Sauv  Meyer, in terms of ‘retrospective’ and ‘prospective’ ways of justifying praise and blame.²⁷ This way of framing the contrast between S-reactive and ‘objective’ attitudes is thoroughly warranted by Strawson’s own view, but it is misleading. As is clear from Protagoras’ celebrated theory of punishing for prospective reasons (Plato’s *Protagoras* 324a–c), *prospectively* praising and blaming an agent S for an action *x* are practices that seek: (1) to encourage S to perform, or to deter S from performing, actions relevantly similar to *x* in the future; or, alternatively, (2) to increase or decrease the probability of the performance of the same type of 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 carries out actions of the type in question, e.g. to warn prospective employers. I think this definition accurately captures the essence of forward-looking or prospective praising and blaming.²⁸ It does not capture, however, the essence of ‘objective’ or what I prefer to call ‘instrumentalist’ attitudes, simply because not all forward-looking attitudes are instrumentalist or ‘objective’.

As Strawson’s definition shows, the distinguishing feature of the ‘objective’ attitude is its instrumentalist character: its being a mere instrument for ‘regulating behaviour *in socially desirable ways*’, an ‘instrument of policy’ that offends ‘the humanity of the offender himself’.²⁹ I will argue in the next section that many cases falling under (1) *supra*, however, do not

²⁶ Sauv  Meyer 1993: 40 (emphasis added).

²⁷ That Sauv  Meyer takes the ‘retrospective’/‘prospective’ distinction to be equivalent to Strawson’s ‘reactive’/‘objective’ is clear from her earlier comments on the moral status of children: ‘In the case of children, we may do so [i.e. praise and blame them] because we expect that they will become morally responsible agents, or because we think that they have acquired some degree of responsibility. But to the extent that we do not think of such agents as morally responsible, we take what Strawson calls the “objective attitude” towards them, and suspend the reactive interpersonal attitudes.’ *Ibid.* 18.

²⁸ Notice that on the prospective view, the objects of praise and blame are (if only as a matter of contingent fact) also voluntary actions. The standard consequentialist argument in this regard, to be found in Stevenson 1944 and Smart 1961, is that praise and blame become efficacious social tools of behavioural manipulation only when targeted at actions that can be ascribed to the agent (i.e. ‘avoidable’ actions, in Stevenson’s sense, or ‘voluntary’ actions).

²⁹ That is, if the offender is a member of the moral community. Strawson 1962: 89 (emphasis added).

satisfy this description in that they are not justified purely by instrumentalist considerations. Moreover, it can be noticed straight away that there are cases falling under (3) that do not do so either. For instance, in warning a prospective employer about S, I am not necessarily treating S in an ‘instrumentalist’ way. Therefore, our subsequent discussion will greatly benefit if we first of all refer to Strawson’s ‘objective’ attitudes as ‘instrumentalist’, for it is the fact that they are purely justified by considerations of social utility that grounds our unwillingness to classify them as ‘moral’ (i.e. ‘moral’ praise and blame); and finally, if we bear in mind that ‘prospective’ or ‘forward-looking’ is much broader than ‘instrumentalist’.³⁰

It is important to note that the onus is on the Strawsonian interpreter to show that Aristotelian praise and blame are not *prospective* attitudes, and that their recipients are not the recipients of prospective attitudes, given that a prospective interpretation of Aristotle has indeed been offered; an interpretation that appears to the Strawsonian as incompatible with his own, and that can offer a perfectly plausible account of Aristotle’s attribution of voluntariness to children and animals, 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’.³¹ Roberts’ own view – let us call it the ‘prospective conditioning view’ – is precisely that prospective praise and blame 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.³²

³⁰ For quite different reasons, the term ‘forward-looking’ (coined by Stevenson) is misleading, in that it wrongly suggests 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 *future response* from the agent targeted. That this is not a distinctive mark of ‘forward-looking’ praise and blame is shown by the fact that 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 apologise, to regret, to make an indemnification, or at least to defend himself against the charge (these are all *future responses*).

³¹ Roberts 1989: 26. For a similar view, see also Frede 2011: 98.

³² Roberts 1989: 25 (emphasis added). In addition to 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.* 23).

Since prospective praise and blame can be used to redirect the future behaviour of children and animals, children and animals are in this sense the ‘appropriate’ targets of Aristotelian praise and blame. Notice that Roberts includes a variety of attitudes such as training, habituation and education (the 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 moral responsibility. I take it that whatever precisely moral responsibility is, animals and small children do not have it.’³³ 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 oppose firmly the prospective conditioning interpretation.

The most elaborate attempt in this direction is Sauv  Meyer’s. She tries to show that the sort of praise and blame that in Aristotle’s view is only bestowed upon voluntary actions, is not even 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 the sort of praise and blame that is 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*. Basing her argument on an analysis of certain passages (*EN* 1101b14–16, 1101b31–2; *EE* 1219b8–9; *MM* 1183b26–7),³⁴ Sauv  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

³³ *Ibid.* 25.

³⁴ For my analysis of these passages, see [Chapter 2](#), section C.

or intemperate, and when he does justice he is properly praised as just. These evaluations are recognizably moral evaluations.³⁵

In saying that an agent who acts the coward or 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 ethically significant actions.

The idea that Aristotelian praise and blame are retrospective, and that therefore these are moral attitudes, and that therefore the sort of responsibility grounded on them is of a moral nature, is of course rooted in the misleading 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.

C. SETTING THE STAGE

We have now identified three distinguishing features of the Strawsonian interpretation. The third feature of this interpretation (i.e. the need to show that Aristotelian praise and blame are not prospective) will be the main focus of the arguments in the next two sections. Nevertheless, before presenting these arguments it is important to note that 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. She thinks that '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*' (emphasis added). But how is this causal sense of 'retrospective' supposed to help us understand why calling an adult agent in possession of an ethical disposition 'just' or 'cowardly' for being the voluntary cause 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*'? How can a merely causal sense of 'retrospective' praise and blame entail a normative, non-tautological 'moral' sense of 'retrospective' praise and blame? It is clear that it cannot.

³⁵ Sauvé Meyer 1993: 51 (emphasis added).

Sauvé Meyer's choice of words to describe prospective praise and blame (i.e. 'tools of behavioural control', 'controlling or altering', etc.) makes it reasonable to think that what she finds objectionable about prospective attitudes is (just as it is for Strawson) their purely instrumentalist character. It is then reasonable to assume that, on her view, retrospective attitudes are exclusively moral in virtue of the following criterion:

(A) If (Aristotelian) praise and blame are retrospective then they are the sort of attitudes that acknowledge their recipient as a moral agent in his own right, or as an end in itself, as opposed to regarding him as a mere *means* or *instrument* (to use a Kantian dichotomy), as instrumentalist attitudes do.

As noted above, Strawson thought that the adjective 'moral' should be withheld from instrumentalist attitudes due to their being mere efficacious tools for '*regulating* behaviour in socially desirable ways', and '*instruments* of policy' that offend 'the humanity of the offender himself'.³⁶ 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 us call this 'the instrumentality argument'.

In section D below, I argue that criterion (A) fails to provide an adequate basis for regarding retrospective praise and blame as *eo ipso* exclusively moral. For the sake of exhaustiveness, however, it is also worth considering whether the Strawsonian interpreter could rely on a different criterion:

(B) If (Aristotelian) praise and blame are retrospective, then they are attitudes that *look back* to the agent as someone who 'deserves' or 'merits' such an attitude in virtue of his being the sort of agent who can be subject to the demands of morality and in virtue of an action of his that has breached or fulfilled a particular moral demand (e.g. Strawson's basic moral demand of goodwill).

For instance, Sauvé Meyer claims that while we may praise and blame (or reward and punish) non-human animals, small children and the mentally defective 'with a view to controlling or altering their behaviour, we do not do so because we think they *merit* such treatment'.³⁷ I argue in section D below (and also in section F below) that this criterion does not hold true of Aristotelian praise and blame either.

³⁶ Strawson 1962: 89 (emphasis added).

³⁷ Sauvé Meyer 1993: 18. A greater emphasis on merit can be detected in the first page of Sauvé Meyer 1998.

I do not wish to beg any questions about whether these two criteria might turn out to be the same, but they should not be conflated, at least at the outset. At any rate, it will be useful to keep them separate for the sake of clarity. Both correspond to the Strawsonian criteria underlying the application of 'moral' to 'praise' and 'blame' (or Sauv  Meyer's 'evaluation'). The crucial point is that Sauv  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). Furthermore, as I am about to show in the next two sections, (A) and (B) *cannot* provide adequate grounds for the Strawsonian interpretation: criterion (A) is not an adequate reason for regarding retrospective praise and blame as exclusively moral, and criterion (B) cannot possibly be fulfilled by Aristotelian praise and blame. As a result, the Strawsonian interpretation cannot even get off the ground. The arguments offered in these sections will also cast serious doubts on the other two features of the Strawsonian interpretation, that is, the view of Aristotelian praise and blame as S-reactive attitudes and the exclusive focus on the morally accountable agent. These other features will receive separate discussion in sections F, G and H below.

D. PRAISE AND BLAME AS INVOLVED IN MORAL EDUCATION

Let me introduce the argument of this section by calling into question Sauv  Meyer's contention that, according to Aristotle, children are excluded from the category of moral agents *for the same reasons* as animals and psychopaths are, namely, because they do not have 'the capacity of reason'.³⁸ Let us assume with Sauv  Meyer that the reason for excluding a given organism from the category of moral agents is the same as the reason for withholding virtue- and vice-predicates from it. If this is the case, then the following passage of the *Ethica Nicomachea* explains why non-human animals are not moral agents:

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 (*ou echei*) decision (*prohairesis*) or reasoning (*logismos*), but rather are outside of <this> nature just like the mad among human beings.³⁹ (EN 1149b31–1150a1)

Commenting on this passage, however, Sauv  Meyer says: 'The reason Aristotle gives here for excluding non-human animals and the insane from

³⁸ Sauv  Meyer 1993: 23.

³⁹ *Ibid.* 23. Her translation.

virtue and vice of character *is also a reason for him to exclude children from these categories of moral evaluation.*⁴⁰ Notice, however, that Aristotle does not mention children in this passage. This seems to me to be of great significance. Aristotle's claim here concerns the *possession* (*echein*) of the capacity of decision and reasoning. While the possession of this capacity is denied to animals, it is not denied to children. Something similar can be said with regard to the next passage Sauvé Meyer quotes in support of her assertion:

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)

Underlying Aristotle's claim that reason is *present* (*enestin*) only in adult human beings, there is an important distinction between 'being present' (*eneinai*) and 'possessing' (*echein*). According to Aristotle, all human beings, *by nature, possess reason* (*echein logon*), at least in the sense that if their 'development is permitted and not maimed, reason *will be present* (*enestai*)' (*EE* 1224b29–30). Children then *possess* reason, albeit in a potential or perhaps merely passive way (the latter being the emphasis in *EN* 1102b31–2).⁴¹ By contrast, here Aristotle is using the verb '*eneinai*' ('to be present in') to characterise the possession of a fully developed capacity of reason as exhibited by rational adults: only in them is reason *present*. This distinction between 'possessing' and 'being present' suggests that Aristotle's claim, in *EE* 1224a27–30, that in children reason is *not present*, is fully compatible with there being a clear sense in which they *possess* reason (e.g. potentially). The mention of animals, along with children, is merely due to the fact that we do not attribute 'conduct' (*praxis*) to animals either.⁴² The fact remains that, in children, reason is not present because it is not fully actual or developed, whereas in animals, it is not present because it is extraneous to their nature. This suggests that the reason for withholding virtue/vice-predicates from an organism is the latter (i.e. lack of possession), and not the former (i.e. lack of presence). Indeed, this is presumably the reason why Aristotle never says of children that they cannot be called 'temperate' or 'intemperate'.

⁴⁰ *Ibid.* 23 (emphasis added).

⁴¹ As Aristotle says in *Pol.* 1260a14, the child *has* (*echei*) the deliberative part of the soul, but in an undeveloped form (*atelês*).

⁴² Not even the *Ethica Eudemia* claim that we do not attribute conduct (*praxis*) to children is as conclusive as it may seem. The *Ethica Nicomachea* and the *Magna Moralia* for instance only deny conduct to animals (*EN* 1139a20; *MM* 1187b7–9).

A brief digression is to the point here. When Sauvé Meyer says that the reason Aristotle offers ‘for excluding non-human animals and the insane *from virtue and vice of character*’ is also a reason for him to exclude children *from these categories of moral evaluation*’ (emphasis added), she is conflating ‘excluding S from virtue and vice of character’ with ‘excluding S from the categories of moral evaluation’. What is relevant for our discussion is the latter. Although Sauvé Meyer does not make this explicit, the conflation in question is based on the assumption that we can apply a virtue/vice-term to S only if S is in possession of the corresponding virtuous or vicious character trait, but this assumption is unjustified, as we shall see in [Chapter 2](#), section C. For the time being, let us just note that the central message Aristotle conveys in *EN* 1149b31–1150a1 is that we do not *call* (*legomen*) animals ‘temperate’ or ‘intemperate’, and that he never says this of children. Sauvé Meyer goes on to note that Aristotle is committed to deny that children *are* virtuous or vicious,⁴³ and she is certainly right, but unless the assumption just indicated receives independent support, it does not follow from this that we cannot apply virtue/vice-terms to them (i.e. that they are excluded from the categories of moral evaluation). Therefore, nothing is gained solely by appealing to the fact that in Aristotle’s view, children are not virtuous or vicious.

Now, 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 (and for Sauvé Meyer) 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, putative exempting conditions such as childhood or psychopathy are all, *for the same reason*, conditions that invite us to regard the agents in these conditions from the point of view of the ‘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’.⁴⁴

Now, in parallel with, and connected to, the Strawsonian blending of these importantly different exempting considerations, the Strawsonian tradition ignores a fundamental distinction between *moral education*, teaching, and the sort of attitudes towards pupils involved in the formation of

⁴³ Sauvé Meyer 1993: 23–4.

⁴⁴ Strawson 1962: 78–9.

their character, on the one hand, and behavioural *control* and *conditioning* on the other, which apply to psychopaths, animals and children alike. The Strawsonian omission of this important distinction is mirrored by Sauvé Meyer 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*',⁴⁵ thus merging two importantly different attitudes, namely, plain behavioural manipulation and conditioning, and moral education ('character formation'). As I have already suggested, the phrase 'prospective attitude' or 'prospective consideration' fails to capture the sense in which Strawsonian *instrumentalist* attitudes are non-moral, because not all prospective attitudes are *ipso facto* instrumentalist. Roberts' prospective conditioning interpretation makes a parallel mistake, when he says 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*.'⁴⁶ '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', and neither animals nor psychopaths are 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 someone is a way of treating him as a rational being in his own right, whereas conditioning him bypasses his rationality. An additional consideration may be that teaching is *for the good of the learner*, whereas conditioning him (if he is a rational agent) is for some good that cannot be *his own*.⁴⁷ We could merge both considerations by saying that

⁴⁵ Sauvé Meyer 1993: 40 (emphasis added).

⁴⁶ Roberts 1989: 23 (emphasis added).

⁴⁷ As I see it, 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'.

teaching is for the sake of the learner's *own good as a rational being*, or put in Aristotelian terms, for the sake of his own ultimate *aretê* (virtue) and *eudaimonia* (human flourishing). Conditioning and controlling someone's behaviour, on the other hand, involves no consideration for the learner's own good as a rational being; it is, in this sense, purely instrumental. If these considerations (or something like them) are correct, moral formation is thus not 'instrumentalist' in the sense in which treating someone instrumentally seems to us so at odds with the concerns of morality (for either of the two reasons mentioned). 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 alternatively (when it is not justified) 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, namely, because they bypass their targets' rationality and treat them in a merely instrumental way, as means to ends from which only external agents benefit, then this reason(s) (i.e. the instrumentality argument) 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', or at least not in the sense in which 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 (a) the rational, mature prohairesis agent in possession of an ethical disposition,⁴⁸ and (b) the immature agent who has not yet developed an ethical disposition; rather than between (a*) what is *potentially* rational – both (a) and (b) – 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 will develop this idea further in [Chapter 2](#).

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,

⁴⁸ I am including continent and incontinent agents in this category, as Sauvé Meyer does.

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’ (viz. his rationality). 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 the parent is doing it for his child’s *own rational good* or for the sake of the child’s ultimate *aretê* (virtue) and *eudaimonia* (human flourishing). The parent is helping him to *become* an ‘autonomous Aristotelian 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 and/or bypass the rationality of their targets.

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, exempting 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. Nevertheless, if we have indeed educational attitudes to children that can be plausibly characterised as ‘moral’, then the Strawsonian reasoning above can be shown to be question-begging. Part of what is essential about educational attitudes is that the sort of blame exemplified by the father’s scolding of 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 the child has acquired full membership of the moral community. Aristotle would agree with this (Chapter 2, section E). 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’).⁴⁹ 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 of 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

⁴⁹ For instance, in the way Aristotle thinks ‘most children’s games should be *mimêseis* of the serious occupations of later life’ (*Pol.* 1336a34).

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.⁵⁰

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, but rather that they are not moral because they are instrumentalist; and I have offered (I think) convincing arguments against this idea. If I am right, these arguments undermine one of the main motivations for trying to find in Aristotle's theory of voluntariness 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 responses 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 first establish 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 [Chapter 2](#), section E.

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).⁵¹ This is, as we noted in section B above, the

⁵⁰ Moreover, it is not even clear that the Strawsonian himself can *consistently* claim that immaturity is the sort of exempting condition 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 pictures the treatment of children involved in their moral education and character formation as one that gradually, though 'insensibly', shifts between objectivity of attitude and genuine reactive attitude. See Strawson 1962: 88.

⁵¹ 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'. Sauvé Meyer 1993: 18 (emphasis added).

other alternative open to the Strawsonian interpreter. In the next section I argue that this alternative is also closed.

E. 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.⁵² Moreover, had such a distinction been drawn by the time Aristotle was writing his ethical treatises, 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 concerning 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 view, is a feature of it that connects it primarily with its possessor: it is in the possession and, more conspicuously, in the exercise of excellence of character, that human beings attain the perfection appropriate to their own condition *qua* human beings.⁵³

There are many examples of what I call 'self-regarding' ethical *hexeis*, that is to say, 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, I

⁵² For instance, R. B. Brandt 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' (Brandt 1970: 24 (emphasis added)). See also Adkins 1960: 336–40, 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.

⁵³ 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).

think, are good examples on the side of vice, and Aristotle himself is well aware of this self-regarding aspect in the case of vices corresponding to the virtue of munificence, like being vulgar 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 (*blaberaî*) to one's neighbour nor excessively discreditable. (EN 1123a32–3)

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

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* (*blaberaî, 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 (praiseworthy) and which ones faulty (blameworthy). This conviction of Aristotle's is relevant to our argument in the following way. By 'moral demands' Strawson and the Strawsonian theorist understand, *inter alia*, the sort of demands the breaching of, or the acquiescence in which makes appropriate or *justifies* the corresponding S-reactive responses. 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 gluttonous person. 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's aforesaid conviction amounts to, namely, that current social practices of praising and blaming should not be taken as a criterion for determining which dispositions are *praisable* and which ones *blameable*, 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 both 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' or 'ways of treating' that look back (as the S-reactive attitudes do) to the faulty agent as someone who deserves or merits them in virtue of an action of his that has breached (or fulfilled) a *moral demand*, in the Strawsonian sense of 'moral demand'. Incidentally, notice that existing social responses of reproach, as reported by Aristotle, are typically aimed at those traits of character that are directly harmful to other people, making it reasonable to think that they are typically the manifestation of S-reactive attitudes like resentment or moral indignation. If an ethical disposition can be considered as a vice and as something blameable regardless of whether it attracts this sort of reactive responses or not, then the sort of blame associated with vice cannot be analysed in terms of S-reactive attitudes like resentment or indignation (notice also that Aristotelian praise and blame may not even be 'attitudes' towards people or 'ways of treating' to begin with; I discuss this in the next section).

As a result, it is fair to say (granting the argument in section D above) that Aristotelian praise and blame cannot be shown to be moral (in a sense that does not merely translate '*êthikos*') by 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 behaviour that has breached or complied with a moral demand (in a non-tautological sense). Moreover, the second feature of the Strawsonian interpretation (i.e. the concern with the moral agent) is, once more, inevitably thrown into doubt. For it is now clear that the Aristotelian virtuous or faulty agent is *not* the Strawsonian, accountable agent, even if the concepts of the ethical agent and the accountable or answerable agent are in fact co-extensive.

F. ARISTOTELIAN PRAISE AND BLAME AS LOGOI

Let me now reinforce my argument by directly addressing the question of whether the praise and blame Aristotle has in mind are S-reactive attitudes. That they are not is clearly shown by Aristotle's own explicit account of praise and blame. For Aristotle, praise and blame are primarily manifested in language itself. This allows him to say that, in a sense, to be ambitious

in the right manner or to be pusillanimous are ‘praisable’ and ‘blameable’ aspects respectively, without implying that they are dispositions that are, or should also be, open to S-reactive attitudes like reactive social esteem or condemnation.

Let me explain. Aristotle repeatedly represents virtue and vice as the ‘formal objects’ of praise and blame respectively: he says that (i) ‘praise is of virtue’ (*ho epainos tês aretês*) and that (ii) virtues are praiseable (*aretai epainetai*),⁵⁴ and even though he seems to use no expression parallel to (i) in the case of vices, he does say that vices are blameable (*‘kakia* are amongst the *psektâ*, *EE* 1223a10). In maintaining that ethical dispositions, virtues and vices are respectively ‘praisable’ and ‘blameable’, and presenting virtue in particular as the ultimate target of praise, Aristotle is clearly referring to virtues and vices as *qualities*.⁵⁵ Now, by these expressions, (i) and (ii), Aristotle certainly does not want to imply that we praise or blame Virtue or Vice personified, as it were. To the extent that virtues and vices are only to be found in individuals, praise and blame *as such* (i.e. as *logoi*) are only appropriately bestowed upon the latter. What he means by expressions like (i) and (ii) is simply that the qualities in question are *valuable* or *worthless* respectively: it is the relationship that individuals have with these qualities that makes them praiseable or blameable respectively (I say more about this in [Chapter 2](#) section C). But there is more to these expressions than this. When we praise or blame an individual S (because he is somehow related to the qualities in question), it is not the case that we first identify his virtues or vices and only then we decide to praise or blame him, as if the act of identifying S’s virtues or vices were different from the act of praising or blaming S. This is absurd, because virtue/vice-terms (‘courageous’, ‘unjust’, etc.) themselves carry a laudatory and pejorative meaning, respectively. By calling S ‘courageous’ or ‘unjust’ we are *already* praising or blaming S.

Aristotle is clearly aware of this semantic aspect of expressions such as (i) and (ii) *supra*. A confirmation of this can be found in his tendency to consider praise and blame as common-sense *tests* for regarding a given trait of character as an excellence or a defect. So, for instance, at *EN* 1103a4–10, when he is arguing for the distinction between intellectual and ethical dispositions, Aristotle 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 due to his disposition, and those dispositions which*

⁵⁴ *EN* 1101b31–2, 1103a10, 1108a14, 1126b65–6; *EE* 1219b8, 1220a5–6, 1223a10–13, 1228b30–1, 1233a4–8; *Rhet.* 1366a37.

⁵⁵ Virtues and vices are *hexeis* (habits or stable dispositions), and *hexeis* are in the category of quality. See *Cat.* 8b25ff.

are praiseable we call excellences'.⁵⁶ If this is partly what Aristotle means by the possibly misleading expression 'praise is of excellence', or by saying that virtues and vices are 'praiseable' and 'blameable' 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 by 'praiseable' and 'blameable' Aristotle cannot mean 'open to social esteem and reproach', at least if these are understood as expressions of S-reactive attitudes (section E above). 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 by this he means that mere terms of affections like 'angry' or 'confident' do not convey a laudatory or pejorative meaning.

Now, praise *sensu stricto*, that is, the praise we bestow on individuals rather than the laudatory meaning of virtue-terms, is unambiguously characterised by Aristotle as a *logos*: 'Praise is the *logos* that exhibits the greatness of <a person's> excellence' (*Rhet.* 1367b27–8); 'praise is a *logos* of the general quality of a person' (*EE* 1219b16).⁵⁷ The closest Aristotle gets to explaining the claim that praise and blame are *logoi* is in *Rhet.* 1367b36–1368a9, when comparing praise with counsel or advice:

Praise and counsels share a common content (*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 (*hupothêkên dunatai*); but expressed thus 'He is proud not of what he owes to fortune, but of

⁵⁶ Also *EN* 1106a6–10 and *MM* 1197a17.

⁵⁷ Here Aristotle follows an old tradition. See Plato's *Laws* 829c, and 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 imitate the virtues of the heroes (236e, 246bff).

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.⁵⁸

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 someone, then you can use the sentence 'he is proud of ...' for such purposes, but the latter sentence 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'. Rather, praise and blame in this sense are what speech act theorists would call 'illocutionary forces' (as distinct from locutionary forces or meanings). They occur as/in speech acts and not as/in the meaning of words apart from the use that we make of them. Aristotle seems to be conscious of this, for he distinguishes '*the things that* you might suggest *in counselling*' ('suggestions') from the grammatical form called 'counsel' (see also *hupothêkên dunatai*, 1367b36).⁵⁹ 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.

This, then, is Aristotle's account of the nature of praise and blame *sensu stricto*: they are just speech acts implemented through sentences in the indicative mood, and containing V-terms (the 'atomic units' of laudatory or pejorative meaning) or descriptions grounding V-terms predicated of one or more individuals, on the occasion of his or their recognisably V-deeds. Notice that Aristotle seems to think that praising and blaming are pre-eminently judgements *in words* (i.e. *lexeis*) of people's or actions' value or disvalue and that expressing these judgements is the same as

⁵⁸ 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 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 view with regard to these issues.

⁵⁹ 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*).

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 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. 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.

Another possible candidate for performing the functions that moral indignation discharges 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 rejoicing 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 neighbours according to their worth – not to the things that our neighbours *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).

G. ARISTOTELIAN PRAISE AND BLAME AND THE ACCOUNTABILITY MODEL

Someone might object at this point that the 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 that, accordingly, the argument in the previous section is too narrowly limited for it 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, in 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, in 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 holding moral agents *responsible*, *accountable* or *answerable*, for breaching, or failing to meet, these moral expectations.

Since the Strawsonian conception may not be the only conception of accountability (even though it is certainly the most influential one), it is worth considering the more general question of whether Aristotle is, or is not, an accountabilist. In order to answer this question, a fitting starting point is to note that a conception of moral responsibility as accountability inevitably places a particular emphasis on our *negative* responses to people's blameworthy actions. In particular, in the case of Strawson 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). The reason for this special emphasis on our negative responses to someone's blameworthy actions is that there is clearly an asymmetry between moral answerability for praiseable actions and the one for blameable actions. In the particular case of the Strawsonian conception, as R. Jay Wallace observes, this asymmetry is explained thus: '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 This is important. Even if the accountability theorist can somehow accommodate the fact that people are also *responsible* for worthy deeds⁶¹ (i.e. for those deeds

⁶⁰ Wallace 1994: 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 recognisable features of accountability. See also Wolf 1990. Ryle 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 (somewhat ad hoc) 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 (Wallace 1994: 70–1).

that fulfil, or even exceed, our moral expectations), whether or not people should also be positively praised for these deeds and credited for them seems rather irrelevant from the point of view of our being answerable to others.

Consider, for instance, the sort of legal accountability one is 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 having breached it. If you fulfil your legal obligations, on the other hand, you are responsible for doing so, in the sense that you are an *accountable subject* (the sort of person who can be held legally answerable for his illegal acts) and your having fulfilled your legal obligations is deemed voluntary (questions about the coercive aspect of the law aside). Nevertheless, what you are doing is fulfilling your legal obligations, that is, doing *what is expected from you*; even if you are not, accordingly, blameable, it is hard to make sense of your being deemed positively *praisable* 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, it follows that accountability sits badly with Aristotle's particular concern with praise, or at least with the fact that for him praiseability and blameability are plainly symmetrical. Clearly, the praise attached to virtue (as its formal object) is partly based on the fact that moral virtue has stature or greatness (*megethos*); that becoming virtuous and excelling oneself is an accomplishment and thereby remarkably hard to achieve (*EN* 1137a9–17, 1109a24–30), and in particular, that virtuous activity normally goes *beyond* what is to be expected (*para to prosêkon*, *Rhet.* 1367b14–15). If the praise (and the special value) attached to virtuous qualities is partly due to these facts, then this is very significant for our present purposes, 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 upon which one can be deemed a good agent, perhaps even respected and trustworthy, but hardly praisable in the sense that Aristotle's virtuous agents and their deeds are.

There is another central aspect of the Strawsonian model, considered as one of our practices of holding people accountable, that is of even greater significance. The negative S-reactive attitudes, which are primary or more basic, can naturally be viewed as dispositions expressed in *sanc-tioning responses*, responses that far from being restricted to moral indignation or resentment, also cover 'avoidance, reproach, scolding, denunciation, remonstrance, and (at the limit) punishment'.⁶² What all these

⁶² Wallace 1994: 54.

unfavourable responses have in common is that they involve the disposition of treating others (to use Gary Watson's apt phrase) 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, and accountability in general, naturally provokes the question 'When is it *fair* to respond in these ways towards an agent on the occasion of such a conduct?' or 'When does he *deserve* an adverse treatment in response to his faulty conduct?'. In short, worries concerning the fairness and desert of a blaming response to an agent arise from the punitive nature of such a response, from its being an adverse or unfavourable reaction.⁶⁴ I explore this further in [Chapter 2](#).

That Aristotle never refers to the notion of *desert* when connecting blame to voluntary actions and their agents is thus extremely significant.⁶⁵ Suppose Aristotle was to say that S *deserves* blame for his voluntary bad action *x*; how would he express such a claim? He would say that S is *axios* ('deserving') of blame in virtue of *x*, or that blame is bestowed upon S *kat' axian* ('in accordance with desert'),⁶⁶ but Aristotle himself never says such a thing. As a matter of fact, there are good reasons to believe that Aristotle's canonical notion of 'desert' or 'worth' (*axia*) cannot even serve to express the normative relationship required by accountability (call it the 'desert-relationship') between blame and voluntary behaviour. Aristotle believes that all things distributed according to worth (*axia*) are external goods

⁶³ Watson 1996: 238. ⁶⁴ On this see Wallace 1994: 4.2–3; and Watson 1996.

⁶⁵ To indicate at this point that Aristotle apparently connects blame (*psogos*) with chastisement (*kolasis*) at the opening of *EN* III (1109b31–5), does not by itself show that blame, 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 first be shown that there is indeed an intimate connection between praise/blame and chastisement/reward at 1109b31–5, 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. That this is surely the case is shown by the fact that the *Ethica Eudemia* 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 (*logô[i] aitias*), debarment in law (*nomô[i] atimias*), and punishment in deed (*ergô[i] zêmias*)'; Demosthenes (*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 axiousti*) than other classes of offenders'; and Plato in the *Gorgias* states that 'if someone has made a mistake *that deserves a flogging* (*ean men ge plêgôn axia êdikêkôs*)' he must submit to it (480c–d). Note the accountability connection between *desert* and *punishment*.

(*EN* 1123b17; *Rhet.* 1387a8–16). Examples of these 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), thereby ruling out blame by definition. The important point here is that considerations of fairness and desert are for Aristotle unimportant when it comes to characterising the relationship between blame and voluntary behaviour, and this is therefore rather strong evidence that he is not concerned with accountability. In particular, it clearly shows that, for him, blame does not express sanctioning dispositions.

Incidentally, it is worth noticing that 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 either, for praise itself is never included in his lists of external goods. This suggests a more general conclusion: when Aristotle says that we praise or blame *S dia x* (*EE* 1219b8–9, *EN* 1101b14–16) or *epi x* (*EN* 1109b30–1, 1110a23, 1127b18), the expressions ‘*dia x*’ and ‘*epi x*’ do not characterise that particular piece of behaviour *x* that makes *S* *deserving of praise and blame*.⁶⁷ The precise meaning of these expressions will be discussed in Chapter 2, section C. Note that I do not mean to imply that it does not matter to Aristotle whether praise or blame 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 the many forms of being treated ‘appropriately’.

H. ARGUMENT FROM SKILLS

There is yet another feature of Aristotelian praise and blame that has not been mentioned, but that is perhaps the most significant of all. It is clear that Aristotelian praise and blame are as pertinent to the domain of skills as they are to excellences and defects of character. As the *Ethica Nicomachea* shows, the application of praise and blame to these two domains, so different from the ‘moral point of view’, is done indiscriminately:

⁶⁷ Talk about desert in this connection is pervasive amongst interpreters of Aristotle. For instance, Sauvé Meyer talks about ‘[t]he sort of praise and blame the agent deserves for (*epi* + dative) these actions and feelings’ (Sauvé Meyer 1993: 51); Pakaluk talks of virtuous people as ‘deserving our admiration and praise’ (Pakaluk 2005: 149). Other interpreters such as Irwin 1980 and Bobzien forthcoming use the term ‘justification’ in this connection (‘justified’ praise and blame), which should be taken to mean the same as ‘desert’ or ‘merit’.

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)

The *Ethica Eudemia* makes a similar point, in a context where Aristotle is seeking common-sense support for his definition of human flourishing (*eudaimonia*) as an activity of a complete life in accordance with perfect excellence: ‘there are praises appropriate to excellence on account of deeds (*hoi epainoi tês aretês dia ta erga*), and encomia appropriate to deeds, and *it is the victorious who are crowned* (*stephanountai hoi nikôntes*), not those who are capable of winning but do not win, and there is the fact that we judge someone’s moral quality from his deeds’ (1219b8–11). These passages strongly suggest that the praise and blame Aristotle has in mind is not peculiar to the domain of the ethical, but rather it also extends to the domain of skills (*technai*).

This feature of Aristotelian praise and blame should not come as a surprise to someone acquainted with Plato’s liberal use of these notions. The reader might recall Socrates’ questions in *Crito* 47b: ‘If a man is an athlete and takes this occupation seriously, does he pay attention to the praise, blame and opinion of every person, or to those of one man only, i.e. his physician or trainer?’, and after Crito’s acquiescence in the latter, ‘Should he fear, therefore, the blame and welcome the praise (*phobeisthai chrê tous psogous kai aspazesthai tous epainous*) of that one man only, and not of the many?’ Later in the *Laws*, in the process of describing the organisation 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).

The bearing of this feature of Aristotelian praise and blame upon our present argument, I contend, lies in the fact that the good runner’s ability to win a difficult race, or the bad musician’s inability to play in tune make them both praiseable and blameable, respectively, in a sense which significantly differs from the sort of praiseability or blameability (in this case, *blameworthiness*) appropriate to agents whose actions we believe have fulfilled or breached moral expectations, and for whom certain forms of moral attitudes like moral admiration or resentment would be appropriate. Of course, judging that someone is a bad musician could involve certain previous expectations or demands, the non-fulfilment of which could explain, for instance, 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 analysed in terms of intrinsically moral attitudes like moral disapprobation or ‘deep’ blame, or any other cousin of resentment.⁶⁸ 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 therefore reflects well or badly upon them as runner and musician, respectively.⁶⁹

I believe the same point could also be expressed in a way that brings out the contrast between Aristotelian praise and blame and the more general notion of accountability. One could say that, according to 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 behaviour. My point could be captured by the claim that, just as we cannot be plausibly regarded as having a right to a good performance from the runner (unless the runner is bound to us by some contract, but that is a different matter), an Aristotelian ethical agent 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, that right is not a

⁶⁸ 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 1990: 40–3; Wallace 1994: 52–3.

⁶⁹ 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 1993: 56, footnote 22 (emphasis added). 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 praiseable in their own right, 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 (*EE* 1219b10) just like the *dromikos* who is, for instance, good at running in the Olympic games, and the *ischuros* who is, for instance, good at wrestling (*EN* 1101b12–18) are in possession of *bodily skills*, sharply distinguished by Aristotle from natural goods at *EN* III 5 1114a21–9, precisely on the grounds that they have been acquired through exercise and care. If Aristotle is not considering skills at all in *EE* VIII 3, then this chapter provides no evidence that, for Aristotle, skills are not praiseable 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).

necessary ingredient in an account of how and why, according to Aristotle, such a performance is attributable to him.⁷⁰

No doubt there are important and interesting disanalogies between the domain of ethical behaviour and the domain of technical performances (for instance, an ethically bad performance is voluntary, whereas a technically bad performance is involuntary, *EN* 1140b22–3), but I cannot see how any of these disanalogies impugn my contention that, according to Aristotle, *the kind* of evaluations appropriate to either domain are simply *epainos* and *psogos*. In general, if the blame bestowed upon ethical agents were, in Aristotle's view, 'moral' blame in a substantive sense, this would be because it carries *in itself* the mark of morality; for example, it would be either S-reactively charged (or involve the belief that being S-reactively charged would be appropriate), or it would be the expression of a disposition to sanction the breaching of distinctively moral demands. Nevertheless, 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 in general.

⁷⁰ I am not assuming that the notion of accountability applies exclusively to actions that have breached *moral* demands. This is certainly not the case: think about cases in which people whose job is to perform a specific task and who are held accountable *to some particular authority* for doing it well or badly. 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 in a race.

Aristotle on ethical ascription
The positive argument

A. ACCOUNTABILITY AND ATTRIBUTABILITY

Admittedly, the Strawsonian analysis of our practice of holding someone responsible (i.e. accountable or answerable) for something *x*, is not the only one available. For the Strawsonian theorist, such practice *fundamentally* consists in holding an other-regarding S-reactive attitude towards S on the occasion of his doing *x*, or believing that such an attitude would be deserved. Other philosophers have maintained, for instance, that holding a person responsible for doing *x* consists in judging that S did or caused something *for which it is fitting that S gives an account*.¹ On this view, what we are doing in holding S responsible is, fundamentally, judging it appropriate that S offers an explanation of his intentions or beliefs about his behaviour, rather than holding a reactive attitude towards S.

For our present purposes, it does not really matter whether the S-reactive attitudes, or sanctioning dispositions more generally, are or are not the basic constituents of our practice of holding people responsible. What does matter, however, is that no plausible conception of accountability can maintain that the sort of blame that ineluctably accompanies the practice of holding someone accountable (even if not the most fundamental element of such practice) *consists* in judging it appropriate that the blamed agent offers an account of his behaviour, for it clearly does not: it still consists in a sanctioning disposition. This caveat is important, because of the connection that Aristotle establishes between voluntariness and our practices of praising and blaming. Because of this connection, any conception of accountability that is of any interest to us as an interpretation of Aristotle's main concern with voluntariness in his *Ethics* will have to include not only the setting of a moral community in which mutually accountable agents require certain moral standards of conduct from each other, but also their

¹ See for example Oshana 1997.

being disposed to respond unfavourably to one another's voluntary failure to meet these shared demands, and their *deserving* these unfavourable responses for thus failing – whether or not the stress is laid upon the mutual requirement for moral standards of conduct, or on the disposition to sanction those who fail to meet them. According to this general notion of accountability, (1) B is responsible to (i.e. accountable or answerable to) A, and to everyone A represents, (2) for complying or not with these moral demands or expectations, and (3) deserving of sanction if he fails to meet them.

The conclusions arrived at in [Chapter 1](#) clearly show 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*, even in the more general sense of 'accountability' just mentioned. In particular, I have argued that features (2) and (3) are not true of Aristotle's conception of what it means for someone to be blameable (and praiseable) for voluntarily doing an ethically significant action. In this section, I wish to argue that there is a different aspect or face of moral responsibility that does not invoke the features associated with accountability, and that correctly captures and illuminates Aristotle's concern with the conditions of voluntariness in his *Ethics*.

Such a model, I believe, is provided by Gary Watson's notion of *attributability* (and in the case of attributing bad conduct, by what Joel Feinberg calls 'imputations of fault').² Watson in particular argues that there is available another aspect 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.³ The interesting idea is that not all cases of being morally responsible for one's actions are analysable in terms of having certain responsibilities or being answerable to other people, or as Feinberg and Strawson would put it, analysable in terms of being

² See in particular Watson 1996. Watson 1987 is also relevant. The distinction between *attributability* and *accountability* also seems to correspond roughly to Feinberg's distinction between *imputations of fault* (being 'to blame' for a fault) and *ascriptions of liability* (being subject to further blaming performances where there are persons 'granted the right or authority by relevant rules to respond overtly and unfavourably to the actor for his registered fault' (Feinberg 1970: 128).

³ It is uninteresting because this is the sense of 'responsibility' involved in the notion of agency, and in this sense an agent can be causally responsible for an involuntary piece of behaviour.

liable to overt unfavourable responses by those who have the authority to respond unfavourably. 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 timid or foolish behaviour also harms others, and thereby violates requirements of interpersonal relations, that is a different matter.⁴

Watson here may wrongly give the impression that *mere attribution* of a certain piece of bad conduct to an agent, as distinct from holding him responsible for it, only makes sense when such conduct manifests a self-regarding vice, that is to say, when assessing such conduct does not require focusing on it in relation to its foreseeable effects on others, or does not imply ‘a public forum’.⁵ But this impression is misguided. What really distinguishes attributability-appraisals from accountability-appraisals is the sort of *perspective* from which they are made. Accountability-appraisals are made from the perspective of someone whose moral interest in an action *x* is delimited by whether *x* (or its agent through it) fulfils a moral expectation or demand of interpersonal relations, and by his entitlement to respond unfavourably if such demand is flouted. Attributability-appraisals are made from the perspective of someone whose moral interest in an action *x* is in turn merely delimited by whether *x* is expressive of the moral quality of the agent (which, by the way, does not have to be a stable character trait), that is, his own adoption of ends; and ultimately delimited by the question of whether these ends are *choice-worthy* or not. Watson calls this perspective of assessment ‘aretaic perspective’ (from ‘*aretê*’, virtue).⁶

⁴ Watson 1996: 231. ⁵ *Ibid.* 230.

⁶ I don’t want to suggest here that it is never appropriate to focus on the moral quality of an agent or his adoption of ends from the accountability perspective. This is certainly not true. To the extent that it is appropriate, however, it is derivative: it is only in so far as the moral quality and the adoption of the ends in question are such *as* to lead to or culminate in an action (or omission) with regard to which it makes sense to ask whether it fulfils or not a demand of interpersonal relationships, that such moral quality and adoption of ends is of any moral interest from the accountability perspective.

There is no reason to doubt that our appraisals of actions from the aretaic perspective should also extend to those actions that have foreseeable effects on others and that can in principle also be assessed from the perspective of accountability. Most cases of interpersonal conduct, however, do not reveal the distinction between these two perspectives of moral assessment; thus the need to concentrate on cases where the conduct that is the object of attribution is ‘nobody else’s business’. But there are exceptions: one needs only to think about an instance of conduct that evidently *is* ‘someone else’s business’, where our intuitions as to the responsibility of the agent persist despite the fact that our dispositions to sanction him have been nullified by the thought that such an unfavourable or adverse response would be unfair. Think about the case 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’.⁷

On the other hand, Watson notes (quite rightly in my understanding) 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”’. Watson correctly observes that 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.’

These remarks by Watson confirm an idea that has been widely held amongst accountability theorists, namely, that considerations about fairness are essential to the perspective from which accountability-appraisals are made.⁸ From this perspective, it is only natural that there should be some response, on the part of those entitled to so respond, to someone’s action in so far as it has breached an expectation regulating interpersonal relations, and it is only natural (if such a response is to serve any purpose at all) that

⁷ Watson 1996: 240.

⁸ *Ibid.* 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 1962: 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 and Ravizza 1998: 187; and Wolf 1990: 85–6.

such a response should be punitive in character, or better, a response that expresses a sanctioning disposition.⁹ Because of the punitive or sanctioning nature of such a response the question of whether its target *deserves* it, or of whether it is *fair* to respond in such a way to his conduct, seems inevitable. For example, the facts about the formative years of the malicious criminal suggest that it would be unfair to respond unfavourably towards his crimes, given that ‘we should not be made to suffer from sanctions which we had no reasonable opportunity to avoid’.¹⁰ The matter seems to be quite different if seen purely from the aretaic perspective, however. Nothing in this perspective requires that blaming judgements like ‘he acted in a cowardly way’ or ‘he acted in an unjust way’ should *per se* involve further sanctioning responses or express sanctioning dispositions.¹¹ To determine why exactly this should be so is not an easy task, and I shall postpone a more detailed discussion of this point until the next section. If the point is granted, however, it follows that the aforesaid judgements are not appropriately characterised as responses that the agent ‘deserves’ (in the narrow sense connected to fairness distinguished in [Chapter 1](#), section G).

Let us now move to Aristotle. That Aristotle adopts the aretaic perspective is conspicuous. In Aristotle’s view, in praising or blaming someone for an action, what is fundamental is precisely whether such an action expresses the agent’s moral quality, which is at least partly characterised by his *adoption of ends*, which are in turn assessed by their choice-worthiness. Aristotle offers a detailed explanation of this point in *EE* II 11. Moral excellence, on his view, is the cause of the end’s being right: ‘because of moral excellence the end will be right (*dia tēn aretēn an orthōn eiē to telos*)’ (1227b35–6; cf. *EN* 1144a7–8, 29–36). Now, for Aristotle, our adoption of a given end and, in particular, the adoption of the end that will regulate one’s life – i.e. one’s conception of the good life or *eudaimonia* – is expressed in our rational choice (*prohairesis*), because our rational choice is ‘of something and for the sake of something (*tinōs kai heneka tinōs*)’ (1227b37). That is, *prohairesis* is always a choice (i) to do an action (ii) for the sake of fulfilling

⁹ The term ‘disposition’ is very important here. Someone could argue that the ethical condemnation of someone beyond reach of sanction, for instance, 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.

¹⁰ Watson 1996: 242.

¹¹ At this point, the Strawson-oriented reader might object that Aristotle apparently connects blame (*psogos*) with chastisement (*kolasis*) at the opening of *EN* III (1109b31–5); presumably, this would be evidence that his concern with the conditions of voluntariness is a concern with accountability. For my response to this objection, see [Chapter 1](#), note 65.

a goal. While moral virtue ‘causes us’ to adopt the right or choice-worthy end (i.e. the right conception of *eudaimonia*), rational choice is a sort of determination, a desire arrived at as a result of rational deliberation, to execute what deliberation has shown to be the best means in our power *to attain that end*. To be acquainted with someone’s rational choice, that is, his practical determination to follow the course of action that deliberation has revealed as the best way of attaining a certain end, is thus tantamount to (i) knowing that person’s conception of the life that is worth living (i.e. what ends he has adopted), but also (ii) knowing his rational ability to attain that goal. Both of these are distinguishing characteristics of a good or bad moral character, strictly speaking (*EN* 1144a6–9, 20–36, 1144b30–1145a6). Presumably, this is why Aristotle says in *EE* II 11 that ‘we praise and blame having in sight rational choice (*eis tēn proairesin blepontes*) rather than acts’ (1228a13–14).¹² Supplemented by Aristotle’s claim that we *do* praise and blame agents *through* (*dia*) their acts (*EE* 1223a11–12), the former amounts to the claim that ethical appraisals are made from the aretaic perspective, that is, from the perspective of someone who regards an agent’s action in so far as it expresses (i) the agent’s own adoption of ends, and (ii) his rational ability to accomplish them.

Talking about the agent’s own adoption of ends might be deceptive, however – and it is important to alert the reader to this. We know that, according to Aristotle, children do not have, strictly speaking, rational choice (*EN* 1111b9), which implies, amongst other things, that they do not, strictly speaking, whole-heartedly adopt practical ends as theirs. But this should not be taken to imply that their actions cannot be correctly assessed from the aretaic perspective. Their actions might not express *their own* adoption of ends, but they do express the adoption of ends of a particular sort of moral agent: Aristotle’s claim that ‘we praise and blame having in sight rational choice (*eis tēn proairesin blepontes*) rather than acts’ (1228a13–14), does not have to mean that we praise and blame having in sight the rational choice *of the agent* whose act has elicited our praise or blame. I will come back to this point in section C below.

We are now in a position to see how Aristotle’s adoption of the aretaic perspective explains the distinguishing features of his conception of praise

¹² This does not mean that actions have a mere surrogate role in praise and blame, for instance, in that we cannot have direct access to the quality of a person’s rational choice and so we are forced to judge him by his actions (*EE* 1228a15–18); this would relegate the role of actions in praise and blame to a mere *evidential* role, but this is not what Watson means when he says that they have an *expressive* role, and it is certainly not Aristotle’s view, for he thinks it is of the very nature of a virtue to realise itself in activity, an activity that is constitutive of *eudaimonia*.

and blame noted in the previous chapter. First, it explains why Aristotle never relates praise, and blame in particular, to certain moral sentiments or attitudes. From the perspective of accountability, blaming someone S for breaching a moral demand is a particular sort of further response to S's action, a response that involves, as Watson puts it, 'a readiness to adverse treatment'. For Aristotle, on the other hand, praising and blaming are just *judgements in words* (i.e. *lexeis*) of people's or actions' value or dis-value, and he clearly believes that expressing these judgements is the same as actually praising or blaming people for their actions (see [Chapter 1](#), section F). This fits nicely with Watson's claim that the aretaic perspective 'seems to collapse any distinction between blaming and judgements of blame'.¹³ Secondly, Aristotle's adoption of the aretaic perspective also explains why Aristotle never refers to the notion of 'desert' or 'merit' when connecting blame (and praise) to voluntary actions and ethical agents, as we noted in [Chapter 1](#), section G. Since blaming in particular does not express a punitive disposition, the question of whether it is fair or unfair to blame someone for performing a bad action does not arise. Thirdly, the adoption of the aretaic perspective also explains why Aristotelian praise and blame are *symmetrical*. Since our interest in an action *x*, from the aretaic perspective, is determined by the question of whether *x* is expressive or not of the agent's moral quality – his conception of *eudaimonia* and derivatively, his practical ability to execute it – both praising an action as an expression of an excellent moral disposition and blaming it as an expression of a deficient moral disposition are on an equal footing. We have already seen ([Chapter 1](#), section G) that this is not the case with accountability, where adverse blaming responses are more fundamental.

Finally, the adoption of the aretaic perspective, I think, also helps to explain why Aristotelian praise and blame are also pertinent to the domain of skills and their lack ([Chapter 1](#), section H). There is nothing in the very idea of attributing an action to a fault or excellence in the agent that requires that such a fault or excellence be of an ethical nature, as opposed to a technical one. If I play the piano clumsily, then I am a bad piano player. If I have acted dishonestly, then I am a dishonest person – my action reveals my dishonest quality. Of course *what* my actions reveal in each case is quite different (in the first case they display me as a clumsy pianist, in the second, as a bad person), and the praise/blame-*logoi* in each case have quite different practical implications. But the difference is not to be found in the very notion of blame itself, which is not to be identified as a distinctively

¹³ Watson 1996: 238.

moral response on the basis of a distinctively moral sentiment like moral indignation or resentment.

Underpinning the aretaic perspective and the moral interest we take in an action from this perspective is the distinctive Socratic concern with living well (*eu zên*), that is, living the best life we can as human beings, the life of *eudaimonia* – and derivatively, the conviction that the exercise of certain moral capacities, the virtues, is an essential ingredient to living well.¹⁴ It is this distinctive Socratic concern that ultimately accounts for the fact that our moral interest in an action (even if in practice indistinguishable from other interests) can focus exclusively on whether it expresses or not the agent's correct conception of *eudaimonia* (or more derivatively, the set of dispositions that account for the correctness of this conception and help us execute it). That Aristotle's concern in his *Ethics* is the Socratic concern, as described above, is unquestionable: 'we are searching for the human good and human flourishing' (EN 1102a15), says Aristotle; and he is convinced that the highest good for human beings, the life of *eudaimonia*, is the sort of well-ingrained and systematic activity that is in accordance with the virtues (EN 1098a7–18; EE 1219a34–9).

The Socratic concern is certainly a moral concern (in a broad sense), and thus the perspective of assessment that it underlies is also moral (in a broad sense). But I will keep the term 'ethical' to distinguish it from the family of moral concerns peculiar to accountability, described by Watson as pertaining 'to social regulation and (more obscurely) to retributive and compensatory justice'.¹⁵ Furthermore, I will call Aristotle's own particular version of attributability, 'ethical ascription'. To the characteristics peculiar to Aristotelian ethical ascription I now turn.

B. PUBLIC BLAME AND ARISTOTLE'S OWN VERSION OF ATTRIBUTABILITY

A consideration of Aristotle's view that praise and blame are *logoi* suggests an interesting revision of modern, 'Watsonian'/'Feinbergian' attributability, 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 and Feinberg, 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

¹⁴ Cf. Plato's *Crito* 48b.

¹⁵ Watson 1996: 243.

appraisal as necessarily involving a disposition to sanctioning behaviour. The difference between Aristotle and a modern 'attributivist' (to coin a term for the occasion) is that the latter would also describe the aforementioned blameable conduct as one that is 'nobody else's business', as Watson and Feinberg do (given that the person on the other end is a sane adult), whereas Aristotle would not describe it as such.¹⁶ A member of a modern liberal democracy may think he is morally obliged to keep his negative appraisal to himself (some remarks of John Stuart 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 purely judgements *in words* or *dictiones* (*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* (the statesman) and the nature of the *polis* (the city-state). The chief end of an Aristotelian *polis* is the good life (*Pol.* 1278b22–4), of which excellence and its exercise is an essential ingredient (1283a25). The best *politeia* (constitution) is the one under which 'anybody whatsoever would be best off and would live in happiness' (1324a23–5), 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* 1099b23–9); hence the description of the *Ethics* as an investigation of a political nature (*politikêtis*, *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.¹⁷

Despite the overwhelming emphasis the ancient Greeks, and Aristotle in particular, 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 theorising

¹⁶ Recall Watson's description of the self-indulgent conduct in his example (section A above) as 'nobody else's business', and his claim that accountability-appraisals imply a public forum. Congenial to Watson's view is Joel Feinberg's belief that only in the case of 'ascriptions of liability' is an overt unfavourable response justified (see note 2 *supra*).

¹⁷ As K. J. Dover notes, public and institutionalised 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 fifth century BC Athens. Dover 1994: 229–30.

about excellence, and not to his being interested only in the ascription of good actions, for he obviously is 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 is something we can infer from what has been said about how to praise (*Rhet.* 1368a8–9, a36–7). 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.¹⁸

But herein lies the problem, and a rather serious one. Even if public ascription of good deeds to good people can be understood in terms of the Feinberg/Watson notion of attributability, it remains difficult to see (at least for us liberal moderns) how it is possible for a *public* and *overt* ascription of bad deeds to a person not to be understood as an expression of sanctioning dispositions (ranging from resentment or moral indignation, to punishment, or the belief that any of these would be appropriate). The difficulty may be more forcefully expressed by bringing the following two assumptions together: (1) first, that the privacy of our blaming judgements, or the thought that we don't have the right or it is not our business to express them overtly, is not an intrinsic feature of attributability (as Watson and Feinberg seem to think), and accordingly, that the failure of pure attribution of bad deeds to express a disposition for adverse treatment cannot be based on the intrinsic privacy of such attribution; and (2) secondly, that whatever purpose (whether regulative, deterring, or even retributive) is served by overtly expressing a blaming judgement, this is only attained in virtue of its expressing a disposition of adverse treatment.

Since I have shown that assumption (1) is probably correct, the next step is clearly to deny (2). There are possibly various ways one could go about challenging the latter assumption, but one of them, I contend, is to take seriously the role of shame (*aidôs*, *aischunê*)¹⁹ in Greek culture, as E. R. Dodds and Bernard Williams have done.²⁰ Although P. F. Strawson includes shame amongst the S-reactive attitudes (in particular, amongst self-regarding reactive ones), it has been suggested that this was a mistake of his. From the point of view of the recipient of blame, the standard

¹⁸ In this regard, Plato's description of praise and blame in the *Laws*, as 'the instruments whereby the citizens are educated individually and rendered more tractable and well-inclined towards the laws which are to be imposed on them' (730b), is particularly telling.

¹⁹ Although both terms are not strict synonyms, Aristotle uses them interchangeably in *EN* 1128b10–35.

²⁰ Dodds 1951: chs. 1 and 2; Williams 1993b: ch. 4.

contrast appealed to is the one between guilt (a genuine self-regarding reactive attitude) and shame. For instance, whereas guilt is said by Stephen Darwall to involve ‘an acknowledgement of one’s blameworthiness that recognises 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 amounting to the adoption of a ‘purely “objective” view of oneself’.²¹ It is noted, for instance, that whereas guilt prompts one to respond in the appropriate ways to someone’s getting indignant at one’s behaviour (e.g. by apologising, making the necessary amendments, or perhaps simply by accepting the sanction), shame ‘inhibits second-personal engagement – one feels like escaping from view’.²² The question then is whether experiencing shame is coupled with an interpersonal reactive attitude, in the way guilt, for instance, is coupled with moral indignation. Certainly, the experience of shame can be a personal response to people’s contempt, derision or avoidance (as Williams suggests),²³ but it is not a response to interpersonal S-reactive attitudes like moral indignation, nor in general does it involve the recognition (as guilt does) that a sanctioning response towards one would be appropriate.

For our present purposes, there are two features of Aristotle’s own account of shame that are particularly important. One is its motivational aspect. In the *Ethica Nicomachea*, Aristotle defines shame as ‘a certain fear of disrepute (*phobos tis adoxias*)’ (*EN* 1128b10–11), and it is more precisely defined in the *Ars Rhetorica* as ‘a certain pain or disturbance with regard to bad things which one takes to be things that bring us into disrepute (*ta eis adoxian phainomena*)’ (1383b12–14). In the same chapter of the *Ars Rhetorica*, Aristotle goes on to explain that the focal sources of shame, the primary things we are and should be ashamed of, are all those voluntary deeds that are due to vice (*hosa apo kakias erga*, 1383b18): actions that manifest cowardice, injustice, intemperance, meanness, boastfulness, etc. (1383b18–1384a8). All the other sources of shame can be explained in terms of this primary focus (cf. *êdê apo kakias mallon*, 1384a14; *apo anandrias gar ê deilias*, 1384a19–20).

The fact that shame is a painful state means that one who *has a sense of shame* (a sense of what would be shameful to do) will try to avoid doing what he takes to be the shameful action. Young people, for instance, even if they often err, are said to be prevented from erring by their quasi-experiencing shame at the prospect of wrongdoing (*EN* 1128b18). Moreover, Aristotle recognises that this will (at least in many cases) provide them with the

²¹ Darwall 2006: 71.

²² *Ibid.* 72. See also Williams 1993b: 88–95 and his Endnote I.

²³ Williams 1993b: 90.

right sort of reasons for action from the ethical point of view. Although retrospective shame (that is, being ashamed, *aischunesthai*, for a past or present shameful act), and being in such a condition as to be ashamed (cf. *to houtôs echein . . . aischunesthai*, *EN* 1128b26–7), are only conditionally good, in that they presuppose that one has done or will do something shameful (*EN* 1128b25–8), Aristotle recognises that doing the right thing from one's sense of shame (*di' aidô*), and thus in order to avoid disrepute or discredit (*adoxia*), is acting from the right motive (unconditionally right, though perhaps not the optimal motive) from the ethical point of view. Disrepute is something that must be feared and it is noble to fear it (*dei phobeisthai kai kalon*, *EN* 1115a12–13), and acting from a sense of shame is acting due to virtue (*EN* 1116a27–8). Furthermore, well-disposed young citizens, because they refrain from base actions due to (prospective) shame (*aperchesthai tôn phaulôn dia to aischron*, *EN* 1179b12), are associated by Aristotle with a moral condition that is noble and truly fond of the fine (1179b8–9).²⁴

The second feature of shame that is of interest to us can be found in the connection between shame and the judgements of other people, those before whom (*pros tinas*) we experience shame (whether retrospective or dispositional). Although Aristotle recognises that sometimes shame will be experienced as a response to the flouting of one's own internalised standards (presumably, at some deeper level of moral development),²⁵ it is important for our present purposes that he sees a conceptual connection between the notion of shame (*aischunê*, *aidôs*) and the idea that there ought to be certain people before whom one experiences shame. This connection is grounded on the important concept of disrepute or discredit (*adoxia*): 'Now, since shame is the impression of *disrepute* (and that for its own sake and not for its consequences) and *no one considers an opinion* (*oudeis de tês doxês phrontizei*) <that is held of us> except *on account of those who form that opinion* (*dia tous doxazontas*), necessarily those before whom people experience shame are those whom they hold in esteem' (*Rhet.* 1384a21–5).

²⁴ My account here slightly differs from the careful analysis of Aristotle's conception of shame (indeed, the best analysis I know of) offered by D. L. Cairns, see Cairns 1993: 414–28. It seems to me that Cairns confuses what I call here the *sense of shame* (or prospective shame) with the quasi-disposition to feel ashamed, because he takes prospective shame to be the 'prospective inhibition of an action one is tempted to perform' (*ibid.* 426, emphasis added). The idea of temptation, however, is absent from Aristotle's analysis.

²⁵ So, for instance, the *Ars Rhetorica*: 'Since shame is what has been defined, it follows that one experiences shame over those evils that seem *either to oneself or to those one respects* to be shameful' (*Rhet.* 1384a15–18).

The conceptual link between the *doxa* (opinion) that is held of us and our *adoxia* (discredit or disrepute) is blurred in the English translation, but it is central to our present purposes.²⁶ According to Aristotle, the experience of retrospective shame (or the disposition to experience it) will be a direct response to the opinion or judgement (*doxa*, *EE* 1233b27; *EN* 1128b24) held of us by people one respects, about one's worthiness; and an opinion of this sort is all we need to make sense of a sort of overt blame that is not the expression of a sanctioning disposition. Since one's credit or reputation varies in accordance with how highly or lowly certain people think of one, the more widespread this opinion is, the more one's credit or reputation is affected. Here is where overt blame comes into the picture: the more overt and public is the attribution of bad deeds to a suitably disposed agent, the more intensely he will experience shame (*Rhet.* 1384a33–1384b1, 1385a8–13); but this is not due to the fact that he will recognise more patently the rightness of the other's grounds for complaint, or recognise more patently that he is to be rightly sanctioned, but rather because it will be *more to his discredit*, or if you wish, because it will increasingly diminish the points he had accumulated. Presumably, this is so even when such an overt attribution is also accompanied by anger, resentment or indignation; but this further punitive element is clearly inessential to the arousal of (retrospective) shame. In fact, Aristotle recognises that satirists and writers of comedy, when targeting someone's failings in their plays – plays where these adverse reactive feelings are clearly absent – can arouse the experience of shame in him (*Rhet.* 1384b9–11).

Presumably, Aristotle's insight is that the experience of retrospective shame (and the disposition to experience it) as a direct response to the public blame elicited by one's failings, is at the basis of the process of moral education of suitably disposed agents, and it eventually gives way at a later stage in their moral development to what I have described as a sense of shame, the inhibitory experience of shame at the prospect of wrongdoing and disrepute that Aristotle is willing to count as a decent ethical motive in the *Ethica Nicomachea*. The details of this process are not my present concern here, however. The point I wish to emphasise now is this: the two features of shame here distinguished evidently show that it is false to assume that whatever purpose is served by overtly expressing a blaming *logos*, this can only be attained in virtue of its expressing a disposition for adverse treatment. The experience of shame, and the sense

²⁶ Also implicit is the ambiguity in the use of *aideomai* plus an accusative referring to a person, which may either convey the idea of someone before whom one experiences shame, or the idea of someone one respects, see Cairns 1993: 2.

of it that one develops as a result, exert a powerful motivational pull in an agent, a motivational pull that can be perfectly legitimate from the ethical point of view; furthermore, the experience of shame that is at the basis of the moral development of suitably disposed agents is not directly responsive to sanctioning dispositions, but rather to the mere opinion expressed in a praising or blaming *lexis*. It follows that whatever purpose is served by overtly and publicly attributing a blameable action to an agent (educational, regulative, etc.), this can be achieved without the need to invoke sanctioning dispositions.

C. 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](#), about Aristotle's concern with the conditions of voluntariness in his *Ethics*. There, we saw that in the *Ethica Eudemia* and *Ethica Nicomachea* passages where Aristotle justifies his concern with these conditions (*EE* 1223a9–15; *EN* 1109b30–4), he makes the following two claims:

- (i) Virtue and vice have to do with (*peri*, *EN* 1109b30; *EE* 1223a14) voluntary responses.
- (ii) Voluntary responses, and not involuntary ones, are praised or 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 'ethical 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 the same in kind (*EN* 1114b26–7).

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 exercise of an excellent disposition. In the *Ethica Nicomachea*, he says 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*)' (1101b31–2). Again, in the *Ethica Eudemia*

he says that ‘excellences and defects and the actions *that spring from them* (*ap’ autôn erga*) are praiseable and blameable respectively’ (1223a9–10), and the *Magna Moralia* corroborates these claims by stating that ‘excellences are praiseable things (*epaineta*): for praise originates from the actions that are *done in accordance with them* (*apo tôn kat’ autas praxeôn ho epainos ginetai*)’ (1183b26–7).

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 F, 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’, ‘blame is of Coriscus’), which is already a way of praising or blaming him: this is praise or blame *sensu stricto*. 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) *supra*. This is suggested by his talking about excellences and faults of character as qualities, as distinct from excellent or faulty individuals like Alexander or Coriscus. But what is the bearing of Aristotle’s point about actions (i.e. praise is of excellence ‘through’ (*dia*) or ‘from’ (*apo*) deeds) upon these meta-claims about praise and blame? The point of these meta-claims is to assert 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*. The parallel claim holds true of vice, *mutatis mutandis*. Now, this is perfectly compatible with the two senses of ‘ethical activity’ distinguished above: their being ‘sources’ of noble and bad deeds,

respectively, can be adequately captured by the claim that virtue-qualities render the agent in possession of them such as to do noble deeds, as well as by the claim that they can function as the ultimate goals (the models or ideals) of moral education and habituation. Something similar applies to vice-qualities, *mutatis mutandis*.

Now, because virtue- or vice-qualities have value or disvalue respectively, or because virtue/vice-terms have laudatory/pejorative meanings (meta-claims), calling (a) an individual like Alexander or Coriscus 'courageous' or 'self-indulgent' on the occasion of his recognisably 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 consideration 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.²⁷ 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 explicitly denies the first alternative 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)

The two claims in the previous paragraph can be nicely captured by recourse to the notion of *focal meaning*, the expression coined by G. E. L. Owen to render Aristotle's technical expression '*pros hen legesthai*',²⁸ 'being called by relation to one case'. This technical expression applies whenever there

²⁷ In fact, when Aristotle recognises 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' (*EN* 1109b16–18), he is recognising 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.

²⁸ Owen 1960.

is some type of relation between various uses of a predicate-term and its primary use, a relation without which such uses could not be understood. An example offered by Aristotle himself in *Met.* IV 2 is the focal meaning of the predicate 'healthy', which 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 is healthy, or that an activity is healthy; but these further predications of 'healthy' are explained in terms of their relation to the focal predication: the apple or the activity is called 'healthy' because it helps to keep the body in a healthy condition, and a given complexion is called '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/vice-terms, and say that in their focal use virtue/vice-terms 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/vice-terms are predicated (A) of immature agents like children or some agents in the process of moral reform. Aristotle seems to have precisely this derivative sense of virtue- and vice-terms when he says, in the *Politica*: 'since the child is incomplete, clearly he 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.* 1260a31–3). In this latter use of '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.²⁹ Moreover, in their derivative use, these evaluative adjectives can be predicated (B) of actions, in virtue of a certain relation that actions have with this focal use. This relation is twofold, according to the *Ars Rhetorica*. Virtue is the *epaineton* ('the praiseable') *par excellence* (*Rhet.* 1366a33–6); accordingly, whatever stands to virtue in the relation of tending towards virtue (*pros aretên*, 1366b25), i.e. the things that produce virtue (*ta poiêtika tês aretês*), must be praiseable (*epaineta*); and whatever stands to virtue in the relation of being its signs or works (*sêmeia ê erga*, 1366b29), i.e. the things that come from virtue (*ta ap' aretês ginomena*), must also be praiseable.³⁰ As Aristotle makes clear, when introducing criterion C *supra*, both types

²⁹ 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 1952: 144. Hare himself refers to Aristotle's remark that 'a good act is the sort of act that a good man would do' (*ibid.* 186).

³⁰ See also *De Virt.* 1249a26–30.

of voluntary action have in common their being such as the ideal agent in possession of an ethical disposition ('the V-agent') would do.³¹

Now, if calling a person or his actions 'just', 'courageous', 'moderate', etc., happens to pick out a fully mature, prohairesis agent and that conduct which is fully expressive of his settled 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 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'.

We have now offered 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). Hence, praising an action as 'just' is like saying 'that's it, that's 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 precisely 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 emphasising the role of action in praise and blame *sensu stricto* (e.g. EE 1223a11–12), the primary force of the preposition '*dia*' is precisely to mark those actions *through which* individuals are praised: i.e. calling an action 'V' (according to criterion C *supra*) is a way of calling its author 'V'.

Let us say, accordingly, 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 *x* as 'V' (on the basis of criterion C *supra*) and [2] praising or blaming a reason-responsive agent S in 'V'-terms through (*dia*) [1]?

³¹ 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 characterised as the actions the repetition of which generates a stable character (b4–5), and these are voluntary actions (EN 1113b26–7; 1114a12–16).

And let me call Aristotle's concern with these conditions, a concern 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 will explain these two notions in the following sections. An action is blameable or praisable (on the basis of criterion C *supra*) only if voluntary. The conditions of praise and blame then include the conditions of voluntariness. But there is more to this; for Aristotle also thinks that praising and blaming an action x as 'V' is a way of praising or blaming its agent S only if S had done x without the relevant sort of pain. This is a separate condition (i.e. [2] in Qe), because x 's being done without the relevant sort of pain is best understood as a condition that is not entailed by x 's voluntariness, as I argue in Chapter 7.

Finally, given that x '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. 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. I will also take for granted, as explained in the Introduction, 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. violence, factual error) nor the ones 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.

D. 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, although they might not be directly observable, they are at least in principle *accessible* to an observer in a variety of ways, like questioning. This is an appealing feature in contrast with the long-lasting Cartesian conception of a voluntary overt action as the physical effect of a mental act of willing or 'volition' to perform the overt action; a conception that was justly criticised

by Gilbert Ryle, partly because mental acts of willing are unverifiable even to the very mind engaged in them.³²

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 *x* as 'V' on the basis of criterion *C*, may seem virtually impossible. The reason for this is that a factual description of *x* alone will not be enough to establish *x* 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 *x* is not enough to establish this, because (i) such characterisation also requires attending to the notion of opportunity or right moment (*ho kairos*, 1104a6), that is to say, the particular circumstances in which *x* occurs, or the particular way (*pôs*) in which *x* 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) the right 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 recognising *x* as 'just' or 'cowardly' is tantamount to recognising *x* as 'what the just or cowardly agent would do', and if recognising *x* under this description requires one to attend to the particular circumstances in which *x* 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 fulfil a very demanding task just in order to start 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 recognising 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 peculiar to ethical ascription. The reason for this is that, according to Aristotle, the proper, impartial judge for all ethical behaviour is the virtuous person. In order to understand this, it is crucial to note that an important feature of Aristotle's doctrine of the mean is that two or more 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. cowardly – courageous – rash) and thus judge of each other on the basis of their own location. So, for instance, Aristotle says that:

³² See Ryle 1949: [ch. 2](#).

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)³³

Aristotle's point here is that faulty characters at the opposite extremes of the same spectrum display a radical disagreement as to how to characterise 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 characterise themselves, Aristotle does not say.)

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. Therefore, 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 1173b23–6 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. The notion of the virtuous or apt spectator replaces the notion of

³³ Rowe's translation, with slight modifications.

authority to make moral demands and respond unfavourably to their being breached, so important to accountability.

E. REASON-RESPONSIVENESS

Let me bring to an end my account of ethical ascription, by saying something about the sort of agents who, on Aristotle's view, are the appropriate recipients of praise and blame. Given that Aristotle characterises praise and blame as *logoi*, that is, speech acts (Chapter 1, section E), it is reasonable to start with the modest assumption that the appropriate recipients of praise and blame are those organisms who are able to absorb the focal message of praise and blame *logoi*, that is, to understand it.³⁴ 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/acting . . .'); and in the case of blame, 'don't act/be like the 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 those agents in possession of 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 (*poiôtês*) of the soul that is capable (*dunamenê*) of following reason in accordance with governing reason (*kata epitaktikon logon*)' (1220b5–7).³⁵ 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 Strawsonian 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,

³⁴ From the mere fact that praise and blame are *logoi*, it does not follow that praise and blame are *addressed* to those who are able to understand them, or that praise and blame are 'second-personal'. Someone may suggest that praise and blame are (in Aristotle's view), 'third-personal' and addressed to persons *other than* the praised or blamed agent (for instance, see the example of praise offered in *Rhet.* 1367b36–1368a9, which is formulated in the third person). If Aristotelian praise and blame were third-personal, however, then the account offered in this section would be deeply wrong. But the analogy between praise and advice, to be found in *Rhet.* 1367b36–1368a9, directly contradicts this suggestion, for advices are formulated in the second-person.

³⁵ I follow here Spengel's conjecture and Ross' excision of '*de*' at b6.

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 organisms 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/vice-terms from animals and insane people: none of these agents have (*echein*) rational choice or reasoning (*EN* 1149b34–5).³⁶ As we saw in [Chapter 1](#), section D, children do have (*echousin*) reason, though this is not fully present in them.

But what exactly is reason-responsiveness? Aristotle's most elaborate account of this notion can be found in *EN* I 13 (1102b13–1103a3).³⁷ In this chapter Aristotle introduces the taxonomy of the soul (*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 (*EN* 1102b13–25): continent and incontinent agents exhibit, in their own peculiar ways, a *conflict* between what reason prescribes and 'something else' that motivates them. 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 metechain*); in other words, to argue

³⁶ 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 1993: 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 generates a stable character (b4–5), (3) and these are obviously actions performed by children or people in the process of moral reform.

³⁷ See also *EE* 1219b26–1220a4; *EN* 1149a25–b2; *Pol.* I 5.

that it is reason-responsive, in spite of its recalcitrant nature. This is crucial 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.

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 interpersonal 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 this eminently interpersonal relationship between an authoritative figure and a subject, as Aristotle sees it, is not a mere metaphorical device, and is much stronger than mere resemblance: it is genetic.³⁸ 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 educative processes, then it is only to be expected that the internal relationship between desire and reason, as exhibited in these developed dispositions, is the internalisation of their interpersonal ancestors (*EN* 1119b13–15, 1138b7–9).

Thus, 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 difference between the non-reason-responsive agent, such as 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*), '[i]n the sense in which we say that someone who listens to or obeys his parents or loved ones

³⁸ As Freud observes, this is precisely the interpersonal relation between father and child, and later in the mature individual the super-ego/ego internal relationship. See Richards and Stratchey 1991: Lecture 31 ('The dissection of the psychological personality').

“has reason” (EN 1102b31–2), and he goes on to illustrate the claim that the non-rational part of the soul is reason-responsive, by citing external, interpersonal practices like admonishing (*nouthetêsis*), reproof (*epitimêsis*) and exhortation (*paraklêsis*), concluding that the desiring part of the soul is ‘something capable of listening (*akoustikon ti*) as if to one’s father’ (1103a3).

In addition to 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 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.³⁹ 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 (1150b32–5).

In EN I 13 Aristotle goes on to distinguish two other stages of reason-responsiveness, this time fully internalised 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 courageous person it is presumably still readier to listen (*euêkoôteron*); for in him it always chimes with (*homophônei*) reason. (EN 1102b25–8)⁴⁰

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

³⁹ That Aristotle in these contexts has in view *rational persuasion* is clearly shown by EN VII 9 (1151b10–16).

⁴⁰ Rowe’s translation with slight modifications.

to be told what to do), however, shows that in the continent the desiring part still has a portion of its original recalcitrant nature. Lastly, full reason-responsiveness is attained when this original, recalcitrant nature of the desiring part has been fully tamed, and, as a result, the desiring part is even readier to listen to, and 'to speak with the same voice' as, reason. This is the case with the fully virtuous agent.

But now, what about vice? If reason-responsiveness was the capacity of the desiring part of the soul to respond *to what appears* as right reason (as distinct from right reason as such), there would be no difficulty in counting definitely faulty agents as reason-responsive. That Aristotle in *EN* I 13 has right reason in view, however, seems difficult to deny. Children, the incontinent, the continent and the fully 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 than the self-indulgent to being persuaded by right reason. The self-indulgent, it is true, has been persuaded or convinced (*pepeistai*, 1151a13) by reason to act self-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' (*EN* 1102b15–16; cf. 1151a22–5, 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 praising and blaming *logoi* of the virtuous spectator are right, but do not appear so to the fully faulty agent. Indeed, the definitely bad person (*homochthêros*) ignores what is wrong and what is right, and his ignorance is constitutive of his badness (*EN* 1110b28–32). If so, then how can he be expected to absorb the message conveyed by praising and blaming *logoi*? The self-indulgent (the paradigm of the definitely faulty agent), 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 always be pursued' (*EN* 1146b21–2); 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 (1151a22–4) and his condition is such that he has lost the ability to form the right opinion about ends guiding his conduct (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 1146a33–4).

These claims are 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 seems 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 apotreptontos oud’ an suneî*); how will it be possible to persuade someone in this condition to change his mind (*metapeisai*)? (EN 1179b26–8)

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

Given Aristotle’s conception of vice, the conclusion that entirely faulty agents are simply not reason-responsive, in the sense of ‘reason-responsive’ that gives content to the notion of being an appropriate recipient of Aristotelian praise and blame, seems inescapable.⁴¹ Since being an appropriate target of Aristotelian praise and blame is, we have assumed, being able to understand and absorb the focal message of praising and blaming *logoi*, and since this ability is grounded on the ability of the desiring part of the soul to respond to right reason, and fully faulty agents have lost this ability at some point in the process of their character-formation, they are not the appropriate targets of praise and blame.

Although Aristotle does not say much about psychopaths and people with abnormal psychological conditions, it is reasonable to think that similar considerations apply to them, with the following difference: while in the case of the bad moral agent any Aristotelian account of his actual loss of reason-responsiveness will have to include the repeated voluntary performance of bad actions as part of an earlier process of habituation

⁴¹ There is a series of passages where, according to some interpreters, Aristotle characterises the fully faulty agent as someone who is afflicted by some sort of regret, self-reproach, and internal conflict (MM 1211a35–b4; EE 1240b12–25; EN 1166b7–25). This view is developed in Brickhouse 2003; Irwin 1988: 379–81; and Irwin 2001. I think, however, that the sort of agent alluded to in these passages is mainly the incontinent agent (or incontinent-like agents like the procrastinator or the lazy).

(as described in *EN* III 5), in the case of the psychopath and the like, the Aristotelian account will most likely be predominantly in terms of his natural, in-born propensities. What about animals? In the case of animals, their soul also has a desiring part (*De An.* 414a32–b7), but this does not imply that it is also reason-responsive within them. The desiring part is not the same in a human being as it is in animals, 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, whenever there is such faculty, and partly defined in terms of it. Animals, however, do not have reason (*De An.* 428a19–24).

F. CONCLUDING REMARKS

Let us conclude Part I of this book by summarising the main argument. We find in Aristotle's discussion a viable conception of moral responsibility, but not of moral responsibility understood in terms of accountability; his concern is, rather, with attributability, and in particular with what I have called 'ethical ascription', and its distinguishing features. In particular, we can see that the praise and blame involved in ethical ascription are neither sanction-oriented, backward-looking reactions focused on what is deserved, nor merely instrumental devices for affecting future behaviour; they are concerned with what can be ascribed to reason-responsive agents (who are to be distinguished from morally accountable agents and agents who are driven solely by pleasure and pain), and have a prospective dimension that is appropriate (as purely instrumental approaches are not) to such beings. I have also established that ascribing an ethically significant action to a reason-responsive agent in terms of praise- and blame-*logoi*, is done from the normative stance of the virtuous spectator, and that (partly as a consequence of this) agents that are fully faulty cannot be counted amongst reason-responsive agents, nor therefore as appropriate recipients of such praise and blame.

We can on this basis understand Aristotle's detailed account of the voluntary and the involuntary as an account of the conditions that defeat the ascription of ethically significant behaviour to such reason-responsive agents. This involves in particular explaining Aristotle's complex conception of violence; explaining his account of coercion and how it figures sometimes as a justification (which does not defeat ethical ascription) and sometimes as an excuse; explaining Aristotle's account of factual error and the implications of culpable error; and the meaning of the 'pain condition' – the requirement that what is 'involuntary' must pain the agent. Examining

Aristotle's account of each and every one of these conditions is the focus of Part II. I expect the reader to see that Aristotle has an original, appealing and very complete account of each and every one of such defeaters, and moreover, that this account is internally consistent, coherent, and compatible with the conception of ethical ascription and its distinguishing features that we have developed in Part I.

PART II

The defeaters of ethical ascription

CHAPTER 3

The definitions of violence

A. ETHICA EUDEMIA 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 (*violencia*, *force majeure*). When one is pushed off a cliff or carried away from the battlefield 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 (1224a9–1225a2), 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.

Since action (*praxis*) is a kind of motion (*kinêsis*, *EE* 1222b29–30), it is plausible to start by treating violence in relation to motion 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 characterise primarily the manner in which motion or rest (and later, action or inaction) occur in the subject of motion,² namely, 'against the

¹ Aristotle sometimes uses the term '*anagkê*' (necessity) as well as '*bia*' (violence) in the context of his discussion of violence. At *EE* 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'.

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' (*paschôn*) of motion (*EN* 1110a2; *EE* 1224a22). In order to preserve this

subject's natural tendency'. The following definition captures Aristotle's initial proposal:

V1: *X's motion or rest is through violence (*bia*[*i*]) if and only if 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 philosophy⁴ 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. Moreover, we 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).

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? A 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 cause⁶ of an element's natural motion towards its natural place

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.

⁵ Against Kenny, who believes that V2 (see *infra*) applies indiscriminately to all cases of violent motion (Kenny 1979: 38).

⁶ As distinct from the formal cause (i.e. the nature of the element), which is evidently internal to the element (as is clear from *Phys.* VIII 4).

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 in its way.⁷

The requirement of an external origin does not figure in the case of violent or forced elemental motion, not because all elemental motion is externally caused, but rather 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 characterise the violence exerted on elemental motion as 'external'. Note also that, because of the absence of an externality condition in VI, 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, he seems to add a second clause, the clause about the external origin:

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

The example given is one of an animal that 'is moved *from the outside*, contrary to its internal impulse' (1224a23), an instance of this latter being

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 1997: 102.

⁸ See also *Pol.* 1254a24–b16.

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 might think that in the case of animals (as distinct from inanimate objects as such), the source of their 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, as we have seen it is not at all clear that the natural motion of inanimate objects (which are *not* self-movers) is not internally generated in the relevant way.⁹ 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 *Ethica Eudemia* 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 into 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' (*EE* 1224a23–5). Clearly non-human animals do not have *logos*, but the general reference to 'animate beings' (*empsycha*) in line 1224a24 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 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. Anger (*thumos*) can also enter into conflict with appetite – recall the example of Leontius in *Republic* 439e–440a – and anger, even though it is usually represented by Plato and Aristotle as being the ally of reason, and as being grounded

⁹ For discussion on this *aporia* about the source of the motion of the simple bodies, see Cohen 1994 and Bodnár's criticisms of Cooper's view in Bodnár 1997.

on its evaluative judgements, is not itself an intrinsically rational impulse; animals can also be moved by it (*EN* 1116b34; cf. *Republic* 441b). Conflict between anger and appetite, therefore, is also possible in non-human animals.

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* 1224a23–5). 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. (*EE* 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 to be internal violence, even if this appearance is later 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. Indeed, notice how carefully Aristotle explains why this is not possible for a non-human animal: 'for it does not possess reason nor *desire that is opposed to it* (*ou gar echei logon kai orexin enantian*)' (EE 1224a26–7). It not clear, for example, 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, and accordingly, genuinely capable of acting against it. 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 below. 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 only if (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* 1224a30–1) that:

(3) All behaviour brought about through violence is painful (*lupêron*).¹⁰

Aristotle, I think, does not mean here to suggest that the behaviour of the simple bodies cannot be characterised as ‘forced’; but he does think, as we shall see, that ‘force’ or ‘violence’ (*bia*) is predicated of elemental motion only ‘because of a certain resemblance’ (*EE* 1224b4–5). Now, there is a clear sense in which (3), the pain condition, follows directly from (2), the contrariety condition, so that (3) is part of V2. Aristotle endorses the commonly held belief that contrariety to appetite (*to para tèn epithumian*) and contrariety to temper (*to para ton thumon*) is painful (*lupêron*) (*EE* 1223a34–5, 1223b20, 1224b16–17) and he clearly thinks that contrariety to reason’s prescriptions, as exhibited in the incontinent, is painful (*EE* 1224b20–1). So the contrariety condition (under any possible interpretation of ‘impulse’) implies the pain condition.¹¹ And conversely, 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 below, 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 – for instance, if 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 clear that the externality condition does not

¹⁰ For the pain condition see *MM* 1188a2–3; *EE* 1223a30–5, 1224a30–1, 36–8, 1224b15–21; *EN* 1110b12; *Met.* 1015a26–33; *Rhet.* 1370a9–16.

¹¹ I say more about this in [Chapter 7](#), section B. Sauvé Meyer would disagree about this, because she says that ‘the downward movement of fire would be forced, but not painful. For this reason pain is best understood as a symptom that the movement in question is contrary to impulse; it is not itself necessary for force’ (Sauvé Meyer 1988: 133). But again: ‘violence’ (*bia*) is predicated of elemental motion only ‘because of a certain resemblance’ (*EE* 1224b4–5), and not strictly speaking.

entail the contrariety condition.¹² This observation will be important for my argument in section F below.

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' (EE 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 continent 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' (b25–9):

V₂*: X's motion or rest is by violence iff (1) its source is in something external to X's whole motivational set¹⁴ and (2) is against X's natural tendency (*hormê*).

Aristotle then makes it clear that if there were not recognisable 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

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 *Magna Moralia*'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 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*, and these problems should 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 (*‘hormê’*) relevant to violence is contrariety to both reason and non-rational desire, or to either of them. In this section I argue that this is 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 concerning the continent and the incontinent agent (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 view¹⁵ argues exactly the opposite: only the incontinent acts involuntarily through violence, because his appetite leads him on without (rationally) persuading him, and (rational) persuasion (*peithô*) is the opposite of violence. This view clearly assumes a restricted notion of pain as the pain or dissatisfaction 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

¹⁵ Corresponding, it seems, to the view assumed in *EN* IIIII24–b3.

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 *Ethica Eudemia* 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 us¹⁷ (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 *involuntary*, 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). Let me explain. Aristotle’s most general argument against these views is that neither the incontinent’s nor the continent’s action x has an external source (in the appropriate sense of ‘external’) being as it is in accordance with non-rational appetite or in accordance with rational wish, respectively. But now, assuming that x is not the result of factual error, this merely shows that x is not involuntary; but x would still be against the natural impulse (i.e. reason, as (B) claims, or non-rational appetite, as (A) 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 (C), which encompasses both views.

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 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 (*EE* 1224b28, cf. *EN* 1111a24–b3), and according to nature (*kata phusin*, *EE* 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 Aristotle: merely

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

¹⁷ On this general use of ‘*eph’ hēmin*’, as it occurs in the *Ethica Eudemia* definitions of the voluntary, see Zanuzzi 2007.

claiming, as Aristotle has done concerning views (A), (B) and (C), that neither the incontinent's nor the continent's action x has an external source (being as it is in accordance with non-rational appetite or in accordance with reason, respectively), and assuming that x is not the result of factual error, would merely show that x is not involuntary; but this is not sufficient to show that x is voluntary, for x would still be painful (either painful because morally wrong, or painful because contrary to non-rational appetite). And the same goes for (C), which encompasses both views.

If the incontinent's and the continent's behaviour is non-involuntary-but-rather-voluntary,¹⁸ as Aristotle wants to say, and if pleasure is the opposite of pain and the mark of voluntariness, as Aristotle often assumes (*Rhet.* 1369b24–7, 1370a10), 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 realisation 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)

Parallel then to the wide notion of natural impulse relevant only 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 Eudemean discussion on violence is brought to an end, leaving us with the notions of 'pain' and 'contrariety to natural impulse' in V2* that are simply assumed, without further argument, 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.

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

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 experiences 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 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. 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 summarised by an example. Suppose a courageous cavalryman is carried away from the battlefield by his horse, which runs away terrified of the violent spectacle. He cannot get out of the saddle, in which he is tightly wedged. Now, the cavalryman, 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

¹⁹ To avoid confusion, suppose that his eating the chocolate bar is the product of violence proper, not a case of coercion.

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).

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 cavalryman is courageous.) When the question is asked to the person riding the horse: '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) 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.²¹

The important conclusion to draw from this is that, given the sort of evaluative description of the cavalryman's behaviour assumed by the blame, e.g. as something recognisably cowardly, his two attitudes towards the battlefield are not equally relevant to the involuntariness of his behaviour thus described. The relevant attitude for the purposes of countering the aforesaid charge is the sort of attitude towards his behaviour *qua* cowardly, which is the cavalryman'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 *ethically significant* pain over other sorts of pain that are irrelevant for the involuntariness of an action recognisably 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 morally 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.*

But why isn't this interpretation of the contrariety condition made more explicit by Aristotle? Well, Aristotle does say that a non-human animal is not subject to strict violence 'because it does not possess reason nor *desire that is opposed to it*' (EE 1224a26–7), which suggests that it is *desire's contrariety to reason* that ultimately fulfils the contrariety condition. But

²¹ In reading this chapter, Timothy Chappell has expressed some reservations about the example, because the Greeks would have regarded our cavalryman as incompetent and a coward *no matter how little* he could do to prevent what happens. This might well be true, but the example is to be understood in terms of Aristotle's own idealised conception of the virtuous person, not in terms of the ancient Greek anthropological conception.

he is not explicit enough, and the reason for this might well be that 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*. In the perfectly virtuous agent, like our cavalryman, 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. The idea is rather that the courageous person does the courageous thing *qua* 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). Accordingly, in our example of the cavalryman, 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),²² and the moral pain our cavalryman experiences at his being dragged away from the battlefield *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 tēn psuchēn* (EN 1166a14). Consequently, if our focus is 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. On the contrary, the idea that 'voluntary' and 'involuntary' are contradictories contributes to leaving the contrariety condition unspecified, because voluntary actions are both actions in accordance with rational impulse as well as those in accordance with non-rational impulse.

It does not follow from these facts about a person's perfectly virtuous 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**.²³ Consider 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 feels a 'rational' dissatisfaction 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

²² The focus on the perfectly virtuous agent, I think, explains the contrariety appealed to in the example of the hand at EE 1224b14, as contrariety both to rational wish *and* appetite.

²³ *Contra* Sauvé Meyer 1988, who thinks that the contrariety in question is contrariety to both reason and desire in harmony (i.e. moral character). This view has the unpalatable result that it makes it impossible for (robust) violence to be exerted on continent and incontinent agents.

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 diabetic person in our example exhibit his divided condition, his continent or incontinent condition. Why not think, therefore, that this agent's *rational* dissatisfaction concerning the overall effects of the externally generated behaviour on his wellbeing, 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 cavalryman'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, recall the question raised at the end of section A above: 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 (*orthos*) 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 (cf. *EE* 1224a26–7); namely, that it is not at all clear that my dog's refraining from grabbing the steak for fear of punishment is 'right' or 'correct', from the dog's own point of view. Now this conflict between appetite 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 that has a *prima facie* claim to be regarded as internal violence (and which, accordingly, should have forced Aristotle to add an externality condition, as is clear from *MM* 1188b9–12).

Only incontinent agents are said by Aristotle, as they were said by ordinary people and philosophers,²⁴ to be 'dominated' or 'controlled' (*krateisthai*, *EN/EE* 1145b36) by something (in their case, by the 'violence'

²⁴ See also Plato's *Protagoras* 352e6–7 and *Republic* 440a.

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).²⁵ The distinction between these two ethical conditions, continence (*egkrateia*) and incontinence (*akrateia*), 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 the ego identifies itself with it. Aristotle makes this explicit at *EN* VIII:

Again, people are called continent or incontinent (*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)²⁶

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 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.²⁷

²⁵ 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).

²⁶ Rowe's translation with slight modifications. See also *EN* 1166a17ff and 1178a2–4.

²⁷ For further support see *EN* 1167b15–16; *Met.* 1015a28 and particularly *MM* 1188b6–11: the version of the contrariety condition offered by the *Magna Moralia* 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 pre-eminence 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 cannot be fulfilled by the incontinent's action. But the adding of the externality condition does not imply that

F. 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 V_2^{**}), and a *core* notion of violence, i.e. a notion of violence merely based on the externality condition. The significance of this distinction will not be fully apparent until Chapter 7. The purpose of this section is to argue that Aristotle is in fact committed to recognising 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 above 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 which S is passively subject to, 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 process), then it must be involuntary. Since S's motion does not fulfil the contrariety condition, however, such motion cannot be involuntary in the sense captured by V_2^{**} . Aristotle must, therefore, recognise a core notion of violence merely based on the externality condition:

V3: X 's behaviour 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).²⁸ 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

the contrariety condition has become irrelevant or needs to be modified: it is still wish or rational choice that is standing for the self of the incontinent.

²⁸ See also *Rhet.* 1369b5–6.

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]*.²⁹ 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 about 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 or not you want this to happen, 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 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 (though a bit clumsily) 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 recognising.³⁰ Now, it is worth pausing over the question

²⁹ *EE* 1226a20–33; *EN* 1112a30–1, b31–2.

³⁰ What about the *Ethica Nicomachea*? At 1110a1–3 Aristotle says that a piece of behaviour is violent (*biaion*) if (1) its origin is external, and (2) is such that the agent or rather the patient 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 *Ethica Nicomachea* about the violence condition *per se*, but it adds two points to the Eudemian discussion. The first one is that it makes explicit that the external source of forced behaviour 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 agent are 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 1951: 97. See also Irwin 2007: 179.) It is evident from Aristotle's language, however, that he believed (2) was a second condition, the non-satisfaction of which does

of why Aristotle does not explicitly draw these conclusions in the *Ethica Eudemia* or in the *Ethica Nicomachea*. 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 attributes to S is a faulty action, and that S did it involuntarily, but also that S did it contrary to his rational inclination, and indeed contrary to the whole of his virtuous character. If Aristotle's focus 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 recognising one.

G. PROBLEM OF HOMONYMY

Now, one could argue that it follows from the three definitions of violence that we have mentioned 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 recognising yet another notion of violence, V₃, which is 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.³¹

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 a project) of finding a suitably general definition of violence that applies to

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 *Ethica Nicomachea* (1110b12, 1111a25). If the non-contribution condition entails the occurrence of pain, then to say that the subject *mēden 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.

³¹ For Aristotle's conception of homonymy see *Cat.* 1a1–16.

all natural beings: after all, Aristotle notes that ‘violence’ (*bia*) is predicated of elemental motion ‘because of a certain resemblance’ (EE 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’ (EE 1224a19). Perhaps then one should concentrate one’s efforts on finding a notion of violence that is the counterpart of *voluntary* motion. Merely focusing on voluntary motion, however, is not enough; for V₁ applies also to animals, and animals do move voluntarily (even if derivatively). What Aristotle needs is rather to discard V₁ as irrelevant on the grounds that 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, that is to say, reason-responsive agents.³²

Nevertheless, the problem of homonymy 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 competing 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.

³² Aristotle confines the application of V₂^{**} to those agents who are capable of *praxis* (EE 1224a28–30) in the narrow sense of *praxis* as ‘moral conduct’, of which children are incapable. 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.

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 ektos*, 1188b19). This is an interesting mark of the *Magna Moralia* discussion: the *Magna Moralia*, like the *Ethica Eudemia*, 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 content to argue 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 *Magna Moralia* 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)³³

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

³³ The translation is MacDowell's.

product of violence because they can be ultimately traced back to an external origin, which Gorgias identifies with the perception 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 (which I number for ease of reference):

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. (*EN* III10b9–15)

The third argument is clearly the most interesting one: it operates under the assumption that the dialectical opponent has³⁴ 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 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,

³⁴ I am suggesting that Aristotle is here (III10b9–15) 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 hêdeia*). The assumed claim at the beginning of the paragraph (b9) is, however, that both *ta hêdeia* 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.

not the external object which stimulates the action'.³⁵ This seems to me, as it seems to Furley, to be just the right response. 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 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* 1113b20–1). 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, rather than the external object which stimulates the action. One way of doing this is by calling into question Gorgias' main assumption, namely, that whether *O* appears as pleasant, as beautiful, etc. to *S* depends solely on features intrinsic to *X* and extrinsic to *S*, that is, features of *O* 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 *O* 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

³⁵ Furley 1977: 49.

³⁶ This is why Sauvé Meyer 1998 correctly argues that Aristotle's view is indeed very close to Stoic compatibilism. Another compatibilist interpretation of Aristotle is Everson 1990. Other very good discussions on this topic are Sorabji 1980: ch. 14, and Natali 2004: ch. 8.

assumption from *EN* III4a31–b21: the way *O* appears to each person is relative to that person's character (*hopoios poth' hekastos esti*, III4a32–b1; *par' auton estin*, III4b17): 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* 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 generalisation holds, there must be a law of nature stating 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 individual's *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. Nonetheless, to reach the desired conclusion Gorgias needs more than this: he needs to argue, not only that (i) *S's perception* of *O qua* alarming is external to *S's* individual 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 moving³⁸ from (i) to the more robust claims

³⁷ I say something more about (ii) in [Chapter 4](#), section D.

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'.

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 alarming 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.

Coercion as justification and excuse I: the Ethica Eudemia

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*, ‘necessity’)¹ is that:

[I] people say (*phasin*) that they did these things (*tauta*) ‘by necessity’ (*anagkas-thentes praxai*). (*EE* 1225a6)

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

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’.

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 *Magna Moralia* goes as far as to say that *to anagkaion* resides ‘in the externals’ (*en tois ekstos*, 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 a range of philosophical and ordinary views that use the terms ‘*bia*’ and ‘*anagkê*’ interchangeably, as 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 distinguishes the two concepts. Because of this, I translate ‘*bia*’ as ‘compulsion’ in a2–17, and then from 1225a17 to a23 as ‘violence’.

why cases of coercion might seem involuntary, however, is best treated as a preconception of coercion to be corrected by further philosophical scrutiny. The Eudemean 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 below) 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 to D below) 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 actions in question (*auto touto*) voluntarily. For it is possible for them not to do the action (*exesti mê poiein*), but to abide the suffering (*hupomeinai to pathos*). (EE 1225a6–8)

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 naturally be taken as representing Aristotle’s own synthetic verdict on the controversy (the passage is numbered for ease of reference):

[III] Or perhaps someone may say that some of these cases are <voluntary> and others are not. For [I] 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

³ Anthony Kenny (Kenny 1979) prefers to interpret ‘*hosa tôn toioutôn*’ as referring to the circumstances or situation of an action, and ‘*mê 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 to the agent who voluntarily got himself into the situation in which coercion takes place (e.g. someone who voluntarily joins a criminal organisation). Kenny’s point is disputable, both exegetically and philosophically. At the philosophical level, the fact that A culpably joined a terrorist organisation (i.e. it depended on him not to join it, and he is reasonably expected to know what joining a terrorist organisation implies) does not imply that, *once* A has culpably joined the terrorist organisation 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) gets into a situation that leads to his ignoring a certain fact, and as a result S does *x*, S does *x* voluntarily (he only gets into the situation 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

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*[?] *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 1225a8–13)

Aristotle divides 'these cases' (*toutôn*, 1225a8) into 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 below). (2) And there are also cases in which 'the action' does not depend on one, that is, cases in which one acts by compulsion in a way, but not without qualification.⁴ (Note that Aristotle has not explicitly described these actions as 'involuntary', but this follows from his claim that 'the action' does not depend on the agent, as EE 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 characterise '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 characterise this

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 (I owe this to Heinaman 1988). Finally, my rejection of Kenny's interpretation does not commit me to the implausible claim that in line a10 '*huparxai*' should be taken as meaning something like the agent doing an action. My view is that '*mê huparxai ê 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).

⁴ As noted in note 2 *supra*, here Aristotle is using '*bia*' in a wide sense. As I suggest in section B below, 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 organisation) intentionally threatening some other definite agent for the purpose of eliciting a definite action' (Fowler 1982: 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.* 332). A second notion is the compulsion notion of coercion. This is

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 track Aristotle’s use of ‘*anagkē*’ 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 he acts voluntarily (i.e. he is blameable), and cases within the same category in which he acts involuntarily (i.e. he is not blameable).

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)

Harry Frankfurt’s view (Frankfurt 1973). Frankfurt believes that psychological compulsion is 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 (a notion which 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.

⁶ I follow Mingay and Walzer 1991 in retaining the *mê bia[i]* in 1225a17.

⁷ At a17 I have punctuated with a period after ‘*praxeî*’, 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 ‘*ē 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 points out in his *Index*, Aristotle often introduces a question with the disjunctive ‘*ē*’ to bring a new suggestion, although in a ‘modest and

In [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:
 - (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 to be cases belonging to the same category (cf. *en toutois*, 1225a13–14) because this category is characterised 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 own 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) the omission of which 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 fix 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 might think that not selling it is even more undesirable, so I fix it and wash it. Hence, one may object to Aristotle's non-substantive concept of coercion that it 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

hesitant' way (see Bonitz 1870: 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 'è ou' is the series of questions in *EN* 1165b14–15, b18, b24.

this in mind and see whether the *Ethica Nicomachea* can improve on this non-substantive notion of coercion (see [Chapter 5](#), section A).

Both (A) and (B) in passage [IV] *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 in fact be 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 *Ethica Eudemia* is forced to identify (1) and (2) in passage [III] *supra* 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. *ETHICA EUDEMIA* 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, 'what I do' can be described in at

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

least three relevant ways (that is, relevant to the ethical assessment of the action):

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 *Ethica Eudemia* 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 *Ethica Eudemia* 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 *Ethica Eudemia* 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 blameable. Quite the opposite, perhaps he should be deemed praisable 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 *Ethica Eudemia* 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]*, 1225a11) 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 *Ethica Eudemia* 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; and yet, his adoption of the end still depends on him in a way that grounds our praising him for this adoption, in so far as it reflects his right conception of the good (although Aristotle does not say this much explicitly). As we have seen, Aristotle usually explains this latter point by saying that virtue is the cause of the end of choice’s being correct (*EE* II 12).

This is why cases of the (2)-type in passage [III] correspond to cases of the (A)-type in [IV] (see section A above). 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 *Ethica Eudemia* 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 his action was ‘necessitated’ and involuntary.

Two questions arise concerning the *Ethica Eudemia* 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 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 *Ethica Eudemia* 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 note 12 below). 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* has not yet been introduced in the *Ethica Eudemia*, and the fact that Aristotle introduces it later (EE II 10) in very elaborate terms strongly suggests that he cannot assume that his audience is already familiar with it. '*Prohaireitai*' at 1225a13 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 below, the relevant question with regard to the 'voluntariness'

⁹ Heinaman 1988: 275.

¹⁰ Besides, it would seem 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'.

¹² Although it is perhaps true that Aristotle recognised something like an action-description of type D2 *supra*, he does not seem to recognise 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 *Ethica Nicomachea*). 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 1978: 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.

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' is meant to convey the idea that it reflects his right conception of the 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 in the second worry mentioned above 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* a given substance *S* is to perform a given function *F*, *S must be* constituted by material conditions *M* of type *T*. 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 substance and the goal dictated by it. Rather, given the independently established nature of *F*, the set of material conditions

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 below).

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 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 artefacts 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? 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 corresponds to the material conditions. What corresponds to the goal or function, on the other hand, is the end of the instrumental action. 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] (section A above) 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 noting 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 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 his daughter, but the fact that he wanted to save his daughter in the first place was not itself ‘necessitated’, and it reflects his conception of the good.

The problem that Aristotle eventually came to address in the *Ethica Nicomachea*, I contend, 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 purely subjective coercion are cases in which the agent is to be blamed for doing the instrumental bad action (e.g. killing someone), because it was not worth doing it given the trifling value of the end (e.g. winning the game); but the *Ethica Eudemia* 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 subjectively coerced agent cannot 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 concepts of lack of necessity and of voluntariness on their 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* involuntarily.¹⁵ One could say that the *Ethica Eudemia* account of objective coercion mistakes the notion of justification for the notion of excuse.¹⁶ As I will suggest in the next chapter, the *Ethica Nicomachea* account of OR-coercion succeeds

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 below Aristotle explicitly chooses to classify cases of coercion under *anagkê* and not under *bia* (*EE* 1225a17).

¹⁵ Austin 1956–7: 8 (footnote 2).

¹⁶ Michael Woods has argued that the Eudemian discussion of coercion can be read in two different ways: as a discussion of coercion as justification or a discussion of coercion as excuse. Woods 1992: 133–4. I think this is wrong: Aristotle distinguishes two types of coercion, one that functions as justification, and the other as excuse.

in doing justice to this distinction, by questioning the *Ethica Eudemia* assumptions about the non-teleological description of 'the action' and its intrinsic badness.

What the *Ethica Eudemia* 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 *Ethica Eudemia* 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], section A above) Aristotle distinguishes two different sorts of exculpatory coercion: (i) one in which the agent acts *by necessity* (i.e. OR-coercion) and (ii) 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, as I shall explain. 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. He defines this notion 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 *Ethica Eudemia* sees it, the OR-coerced agent does the instrumental action 'by necessity' and involuntarily, and it rather suggests that Aristotle is now concerned with a plea of coercion that draws its exculpatory force from the fact that the agent acted on an (objectively) *overwhelming inclination*, not from the fact that his evaluative weighing of the 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. (*EE* 1225a23–5)

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 1225a23–5 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 characterisation of cases of OO-coercion is, I think, much more tenable, for rather than emphasising 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 generalisation) it emphasises 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 *Ethica Eudemia* misleadingly expresses this point by saying that the instrumental action was 'involuntary'). What Aristotle says about OO-coercion in the *Ethica Eudemia*, on the other hand, leads us to suppose that a sufficient basis for condoling with S for doing something *x* that is blameable, 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 in this sort of case are quite different, for at least two reasons. Let me mention the first one here, and the second in the next section.

First, it seems that the pain to be avoided and the corresponding fear can be characterised 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 of something painful and violent, rather than as something *wrong* (1225a23–5). 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 in the same balance, the result is that one can condole with S for doing *x*, whether '*x*' stands for harmlessly revealing a trifling secret or for killing one's 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 noting 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 as a matter of fact Helen did 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 *Ethica Eudemia* view on OO-coercion (i.e. the view that the moral 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’ (*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 analyse ‘*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 or in deliberation 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). Aristotle 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 *C*₁ . . . *C*_{*n*} had been present, *S* would not have done *x*.

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 (*noêsat*) is *ep’ autô[i]* if one thinks whenever one *desires* to think (417b24, cf. 427b17–18).

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, clearly recognised 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 Let us examine this further condition.

That it is the notion of *to 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 1225a30–2: '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 irresistible 'passions and thoughts' started at 1225a19.²⁰ Incidentally, notice that Aristotle here mentions 'thoughts and reasonings' along with affections, *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' (1225a25–6). 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 affections (*pathê*) like love, anger, or fear either, 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] (section C above) 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, it is the pain, rather than the fear of suffering it, that is such that no one's nature could withstand. But surely (1) and (2) are connected, and in what follows the phrase 'overwhelming inclination'

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.

²⁰ See also *EN* 1149b3–6.

should be understood as reflecting this connection. Perhaps Aristotle's idea is that (1) depends on (2), that is, that we recognise an inclination like fear as 'overwhelming' or 'beyond nature' at least partly because we recognise that the pain away from which that inclination is moving us is such that no one's nature could withstand.

So, let us add a second condition to our analysis of 'x depended on S':

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

Now, 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 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) *supra* 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' (*EE* 1225a22), nor 'naturally capable of doing violence to nature' (*EE* 1225a23). It is now time to explain the meaning of these expressions.

It is quite tempting to regard this notion of 'what depends on us' as applied to pains or the corresponding fears (i.e. pains or fears that our basic nature can bear or that everyone'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 psyche of the agent to resist. I don't think this question arises for cases of OO-coercion as conceived by Aristotle, however. That is to say, 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, although 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

²¹ Irwin 1980: 121.

(*phusis*), rather than that of a desire which (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.

The appeal to '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) *supra*. 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). For instance, 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 of 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 *Ethica Eudemia* and the *Ethica Nicomachea* that cases of OO-coercion call for sympathy (EE 1225a22–3; 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,

²² Consider the account of irresistible desire offered by Mele 1992. 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 1992: 86 (emphasis added).

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 recognise 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'. This notion of human nature 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 realise that *he* was, as a matter of fact, unable to resist, but rather because I recognise 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 will 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. The reason for this is that 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 peculiar 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 as such. The Aristotelian virtuous imputer cannot condole with S in the midge

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 pity, *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 (*phorai*) and vices (*lunai*). But these things are *not* pleasant, but only pleasant to *them* and to those in these conditions' (EN 1176a15–22). Ross' translation with slight modifications.

case because he (viz. the imputer) recognises the fear as a manifestation of vice.²⁵ Again, psychological 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-coercion, where the agent's adoption of the end 'depends on him' in that it is not based on an impulse rooted in our basic human nature (i.e. an impulse that no one should be expected to withstand). Suppose, 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) *supra* 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 difficult question of why Aristotle, after claiming that objective coercion does not exhibit the characteristics of violence, *bia* (1225a17), persists in characterising cases of OO-coercion as cases of violence (1225a22–3). 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 previous chapter, Aristotle

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 coerced: '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-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 1986: 212.

understands the externality condition as 'external to the whole soul' (*EE* 1224b24–8), 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 objectively overwhelming desires or passions, which do violence to human nature (or, alternatively, which no good human being should be expected to resist).

Coercion as justification and excuse II: the Ethica Nicomachea

A. *ETHICA NICOMACHEA* 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 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 to cases of OO-coercion, which *EE* 1225a19ff identified as cases in which the agent acts on inclinations that are beyond nature (*hyper 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 *Ethica Eudemia*: 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 polar opposite opinions: ‘it is controversial whether these actions are involuntary or voluntary’ (1110a6–7). Examples are the tyrant who, having someone’s family in his power, gives him 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 *Ethica Eudemia* operated with a non-substantive notion of coercion that included cases of subjective coercion, that is, 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 (*EN* 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 *Ethica Eudemia*, 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 and cases in which one acts because of *what one believes to be* something fine.¹

Moreover, just like the *Ethica Eudemia*, the *Ethica Nicomachea* 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 Chapter 4. In this section and section B I will concentrate on his departure from the Eudemean analysis of OR-coercion. I discuss his view on OO-coercion in section C below.

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' hauta*)' (1110a19).

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 basic dispositional 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*

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

kai theia aretê, EN 1145a19), bestiality (*thêriotês*, EN 1145a30–3) 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. The Greek adverb '*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 H. H. Joachim says. 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. The only way I can make sense of these interpretations is on the basis of J. S. Mill's 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'.⁶ This is not Aristotle's point, however; 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 and his rational plan) patently self-damaging and thus irrational. To make Aristotle's claims in A2 intelligible, I submit, 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 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,

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 1951: 97.

⁵ It is surprising to find interpreters simply accepting this claim as intelligible as it stands. Apart from Joachim, see also Kenny 1979: 31–2, and Klimchuk 2002: 7.

⁶ Feinberg 1986: 125

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 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 tōnde*, 1110b3–4).

We saw in Chapter 4, section A, 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 losing the game). Now, (a) and (b) were deemed insufficient for the purpose of distinguishing coerced actions, in this 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 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-bankruptcy 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 delivered the cargo).⁷ 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 *supra*, the main point is that the *Ethica Nicomachea* seems to agree with the *Ethica Eudemia* in that ‘the coerced action’ when taken without further qualification and for itself, is *involuntary* (EN 1110a18, b5). I take it that this is just another way of saying that provided the captain or the driver are rational agents, and given their rational priorities, they would not throw the cargo overboard or help the criminal escape voluntarily under an *incomplete description* of these actions, a description that does not reveal the new practical situations they are in. The *Ethica Nicomachea*’s most significant 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> (*lekteen*) with reference to the time when one acts (*hote prattei*)’ (EN 1110a15).

So that to say that the *Ethica Nicomachea* ‘agrees’ with the *Ethica Eudemia* in that ‘the coerced action’ in its non-teleological description (and relative to the agent’s rationality) is involuntary, is an overstatement. Strictly speaking, the incomplete, non-teleological description of the action is irrelevant to the question about its voluntariness. The Eudemian assumption about the non-teleological description of ‘the action’ (Chapter 4, section B) is thus false.

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 might 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*

⁷ 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).

en tois kath' hekasta)' (110b7, cf. 1107a32, 1141b16), then it might seem as if 'the action' in 110a13 refers to an action-type or universal and 'actions' in 110b7 refer to action-tokens of that universal, thus suggesting that Aristotle's point here somehow relies on a rather mysterious and unargued metaphysics of action. I don't think this is a feasible way to interpret Aristotle's point here.⁸ 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.⁹ Aristotle's point at 110b7 (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).¹⁰ Since the *telos* is amongst these circumstances, indeed, since it is the *main* circumstance,¹¹ 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:

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*)'. (EN 110a12–13)

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 (*ou probaireitai*) 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 *Ethica Eudemia* 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

⁸ Taylor's translation of 110a13 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?

⁹ *Summa Theologiae*, 1a2ae. 7, 3.

¹⁰ 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 or 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.

¹¹ I further elaborate on this claim in [Chapter 6](#), section B.

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.¹²

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 blameable. 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 *Ethica Eudemia* 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 *Ethica Eudemia* can accommodate 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 characterising 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 *Ethica Nicomachea*, which remedies the deficiency in the *Ethica Eudemia* 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 (as we are about to see) against the

¹² All commentators would disagree with me here, for it has been traditionally thought that the main *Ethica Nicomachea* 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 ones 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 autō[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, 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.

Eudemian assumption of the intrinsic badness of the coerced action, that under the relevant teleological description the action is *not* wrong. 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 *Ethica Nicomachea* 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 *Ethica Eudemia*): (i) helping the criminal escape and (ii) saving his daughter; and that (2) he does them *voluntarily* (as in the *Ethica Nicomachea*), 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 *Ethica Nicomachea* 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. PRINCIPLE OF PREFERENCE

This is the occasion to turn to Aristotle's acknowledgement 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 amongst good things (*en tois agathois*) the removal (*apallagē*) of that which is evil or seems evil, or the exchange of a greater evil for a less (*anti meizonos kakou, elattonos metalēpsin*), because these two things are in a way preferable (*haireta*); in like manner, I reckon among pleasant things (*en tois hēdesin*) the removal of that which is or appears painful, and the exchange of a greater pain for a less. (*Rhet.* 1369b23–8)

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 can 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 *Ethica Eudemia*, 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’*, an effective equivalent of *‘apallagê’* in this context) 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 tōnde haireta)*’ and ‘*now and in exchange for these things (anti tōnde)* 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’) that he considers as eligible (*haireton*), and therefore as good; this 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 *Ethica Eudemia* and the *Ethica Nicomachea*. From the point of view of the *Ethics*, both characterisations 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.¹³

¹³ This question about incommensurability is closely connected to the problem of incontinence (*akrasia*) as Plato and Aristotle view it. On this see Wiggins 1978/79, and Charles 1984: ch. 3.

PP implies that what the OR-coerced agent does (i.e. the exchange itself) is *good*. That the *Ethica Nicomachea* adopts PP is also clear. 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 *Ethica Nicomachea* is particularly insistent upon the fact that paying attention only to this aspect is wrong. But then why say that the victim of the tyrant’s threat *does* something shameful? Is Aristotle unaware of the logical consequences of his view in the *Ethica Nicomachea*?

I think he is not. Certainly, characterising the instrumental action as something ‘shameful’ (*aischron*) seems to have different implications to characterising it as something ‘bad’ (*kakon*, EE 1228a18–9; *phaulon*, EE 1228a14). In particular, in the context of the EN III 1 there is reason to think that the ‘shameful thing’ is not something that the coerced agent *does*, but rather something that he *endures* (*hupomenein*, EN 1110a21, a22, a30). As Karen 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.’¹⁴ If this is right, when 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, therefore, that the *Ethica Nicomachea* successfully isolates the distinguishing feature of objective rational 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

¹⁴ Nielsen 2007: 283. Nielsen is arguing against Stocker’s and Nussbaum’s view that Aristotle in the *Ethica Nicomachea* 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 recognised as such by the hypothetical agent. I think her arguments against this view are very cogent.

cases of coercion (in the non-substantive sense) agents are praised, whereas in others they are blamed, Aristotle says (the passage is numbered for ease of reference):

(1) 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 deî*) because of what is such as to overstrain human nature (*tên anthrôpinên phusin huperteinei*) and such that no one could withstand. (2) But perhaps in some (*enia*) 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* 1110a23–9)

The cases of coercion mentioned in the first assertion are clearly cases of what we have called 'objectively overwhelming coercion'. But what is the connection between these cases and the ones mentioned in the second assertion? Are the cases in (2) a sub-class of the cases of OR-coercion referred to in (1)? It is reasonable to suppose they are (cf. *enia*, a26).

Now, the *Ethica Eudemia* recognised cases of OO-coercion, in which sympathy (*suggnômê*) is bestowed upon the overwhelmed agent because he is subject to an impulse that is 'naturally capable of overstraining nature' (*EE* 1225a22–3), 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 (1) that in cases of OO-coercion the agent deserves sympathy (and thus acts involuntarily), while admitting in (2) 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'). In this section, I intend to answer this question and to show that by noticing the existence of recalcitrant cases in passage (2), the *Ethica Nicomachea* account of OO-coercion does not depart from the one offered by the *Ethica Eudemia*.

One possible retort to this way of posing the question is to bite the bullet and argue that the *EN* 1110a23–9 passage suggests that cases of OO-coercion are in fact *voluntary* despite their being the object of sympathy. K. 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'.¹⁵

¹⁵ Nielsen 2007: 277. See also the commentary in Burnet 1900, and Klimchuk 2002: 8. This also seems to be Nemesius' interpretation, who offers the example of Zeno 'who bit out his tongue and

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 both in the *Ethica Eudemia* and in the *Ethica Nicomachea* 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* 1225a21–3; *EN* 1109b32, 1111a1–2). Secondly, Nielsen's solution assumes an unjustified generalisation, 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 (2) is only about *some* cases (*enia*, 1110a26).

I think it is more plausible to think that *EN* 1110a23–9 agrees with the *Ethica Eudemia* in that the overwhelmed, coerced agent acts *involuntarily*, while also allowing for *limiting* cases – which the *Ethica Eudemia* does not explicitly recognise. 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?

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.¹⁶ 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, for it clearly is not. What he means, I take it, is that the appeal to overwhelming coercion, that is, to a fear

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 1565: 384.

¹⁶ See the commentary on fragment 69 in Nauck 1926. The relevant fragment in the *scholia* referring to the threat reads '*katarasmenou autou, ei mê apokteinê[i], akarpian te gês kai ateknian*'.

or pain 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 the 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 characterised 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 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 *Ethica Nicomachea* denying assumption (ii), i.e. denying the incommensurability of pain and ethical wrongness, and pointing at cases such as 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 cannot really see how the judgement that a given pain or the corresponding fear is humanly unbearable (which is the ground of our condolence) can ultimately depend for its justification 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 sympathise 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 sympathise 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 Alcmeon'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 it in [Chapter 4](#). It is reasonable to conclude, therefore, 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'.

Factual error and the source of blame

A. GENERAL ACCOUNT OF FACTUAL UNAWARENESS

According to *EN* III 1, acting *di' agnoian* is one of the reasons one can offer for saying that one acted involuntarily (*akôn*) (IIIob18, IIIIa17, a22).¹ The most natural rendering of the phrase '*di' agnoian*' is 'through ignorance', but this translation lends itself to confusion. First, 'ignorance' is not precise enough, because acting through this sort of ignorance or unawareness (*agnoia*) is acting through 'unawareness of the particulars' (*he kath' hekasta agnoia*) as opposed to 'unawareness of the universal' (*he kathoulou agnoia*, *EN* IIIob32–3).² Whereas unawareness of particular facts or circumstances (e.g. that this *x* was someone else's book) has exculpating force, ignorance of moral prescriptions (e.g. that it is wrong to take someone else's possession without their consent) has not – on the contrary, it reveals the defective character of the nescient agent (*EN* IIIob27–9).³ Secondly, 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 instance, one of Aristotle's examples is that of Merope (*EN* IIIIa11–12) 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. For

¹ *EN* IIIob18–24 and IIIIa19–21 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 A complete list of the particular circumstances Aristotle has in mind can be found 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 *EN* V 8 1135a24–6, *EE* II 9 1225b2–6 and *MM* 1195a17–19 enumerate. I will show below that Aristotle is inclined to regard conditions (II), (V) and (VI) as identical.

³ Some philosophers have challenged this idea; see in particular Rosen 2003.

these reasons, I will translate the phrase '*di' agnoian*' as 'through factual error'.⁴

Now, Kenny's notion of 'positively mistaken belief' is important, but stands in need of 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).⁵

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 would be extremely unsatisfactory for someone who wanted to know why she killed her own son, this being the description of the action singled out by the blame. The excuse seems incomplete in case (1) because the person was in fact her enemy. There are perhaps two ways in which we can complete the description of Merope's 'doxastic situation' in case (1). Her plea that she was unaware that this person X 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 characterised as a sort of 'positively' mistaken (partially false) belief, but I doubt (1b) can be so characterised.

Case (2) is a clear instance 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 through factual error are like (1a) and (2), it would turn out that all cases of acting through factual error are in fact cases of acting on a positively mistaken belief. But this is not the case. Let me suggest why.

At least some cases of the (1)-type are better accounted for in terms of (1b), i.e. the *absence* of any beliefs as to whether X is F (where 'F' stands for that property of X that is relevant to the blame, e.g. 'her son'), rather

⁴ 'Error' here is not meant to translate Aristotle's notion of *hamartêma* ('mistake'). *Hamartêmata* or mistakes are the *actions* done through factual error, not the factual errors themselves (*EN* 1135b13, 1136a7, 1110b29; *EE* 1246b3, b7).

⁵ 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 her son's supposed death, killed him while he slept. See Nauck 1926: 497–501.

than in terms of (1a). Consider the example of Aeschylus at *EN* IIIIa9–10. If we are to believe Henry Jackson,⁶ 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 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 through factual error 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 (for instance, 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’. The concept of ‘error’ in ‘through factual error’ should be taken to include these two sorts of mistake.

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 or mistake that can be characterised 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 through factual error (*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 disputable examples offered in *EN* III 1: the catapult case (*EN* IIIIa10–11) and the wrestler’s case (IIIIa14–15). The catapult case is the case of a man who wanted to

⁶ Jackson 1901. Jackson cites *Republic* 563 C.

show how the catapult worked, and let it go off. Aristotle offers this as an instance of error about *that on which* he acted (*peri ti ê en tini*), so presumably he thought (as C. C. W. Taylor notes)⁷ that the case was one in which the man 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 man, 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, thus interpreted, the catapult case is significantly different from the rest of the examples of acting through factual error, for it is not obviously a case of epistemic factual error; apparently there is no relevant fact that the man with the catapult was unaware of. The other case is that of the wrestler who, wanting only to grasp his opponent's hands,⁸ 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 through factual error, for the error here is not really an epistemic one, but rather that he didn't 'weigh up his strength', or was *careless*.⁹

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 precondition of his having noticed what he was doing and how. For instance, 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. I will come back to this point later.

Now, these examples involving inadvertence and carelessness seem to show that Aristotle's notion of acting through factual error is even broader than epistemic error or mistaken belief,¹⁰ embracing also cases in which

⁷ Taylor 2006, see commentary.

⁸ '*Apocheirizomenoi*' refers specifically to people who practise a special kind of wrestling in which wrestlers grasp each other's hands and are not supposed to clasp their bodies, or hit each other.

⁹ Kenny also notes this difference, see Kenny 1979: 51.

¹⁰ Perhaps it is due to this that Means 1927 translates '*di' agnoian*' as 'unintelligently'.

some psychological deficiency, or perhaps even a character flaw, 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 of ‘psychological deficiencies’ a variety of dissimilar dispositions, and it is important to draw some distinctions. On the one hand, we have what the nineteenth-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 or inadvertence also belong to this family. I will call these ‘vices of impure agency’,¹² because, 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). Now, Aristotle’s view with regard to the operation of vices of impure agency like negligence, heedlessness, etc. (taken in this dispositional sense) is simply that of an *epistemic error* (i.e. mistaken belief) that can be *attributed to a further fault in the agent*. Aristotle seems to think that in cases of acting through factual error 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. I will argue for this in sections D and E below, so I ask the reader to take this for granted in the interim.

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 as finely controlled as would be required to achieve one’s purpose successfully, and where the

¹¹ See Austin 1873: 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 1990: ch. 7) are vices of impure agency, but not those in which the agent adverts to the possible consequences.

¹² ‘Vices of impure agency’ are the opposites of what Walker 1993 calls the ‘virtues of impure agency’, defined by her 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 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’ (Walker 1993: 241).

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. If these cases pose a challenge to the unity of Aristotle's notion of factual error, this is because in many of them it would be simply false to say that the agent did what he did believing something about what he was doing or, in other words, consciously doing it under some (true) description. If some of Aristotle's examples are of this sort, that is, cases in which there is no mistaken belief to be invoked at all, then it looks as if the notion of acting through factual error has indeed no unity. Take, for instance, a clear example where there is no mistaken belief to be invoked at all: through a careless and unconscious swinging of my arms while walking I drop a precious vase, which I am perfectly aware is there, when trying to get through a narrow corridor. If I may anticipate: Aristotle would say that there is no *agency* involved in this sort of example, if we are to base our conception 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 would be no agency involved.¹³ Since (ii) is conceptually prior to (i) (i.e. factual error presupposes agency) 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. Another example: a native English speaker says 'I like your horse' instead of 'I like your house'. Condition (ii) is not fulfilled in this case either.

I contend, however, that even though some of Aristotle's examples may be correctly described as cases of clumsiness, none of them are such that they do not involve a mistaken belief, as in the example of the vase. The case of the wrestler (under the challenging interpretation), for instance, is not relevantly similar to the case of the vase, for whereas there is no agency involved in the latter, there is agency in the former. To see why, it is important to note that the wrestler's case is better accounted for in terms, not merely of his being unaware of *the way in which* (*pôs*) he was actually doing something, but rather ultimately 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, accordingly, that the

¹³ Compare Davidson's account in his 'Agency', in Davidson 2001.

how condition could be subsumed under the *what* 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 an example the person who ‘thought that one was not *hitting* <someone>’, but does not mention the *how* condition, 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, therefore, the case of the wrestler (under the challenging interpretation) no longer offers 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, in particular, a case of believing that the action amounts to touching B violently, TD₁, and not killing B, TD₂. The case of the catapult (under the problematic interpretation) can be accounted for along the same lines; e.g. while trying to show how the catapult works, the boy points at the firing mechanism with too much emphasis and sets it off, that is to say, he believes that his action amounts to pointing at the firing mechanism with emphasis, TD₁, and not to setting it off, TD₂. These cases, therefore, do involve agency. In the next section I will return to these considerations concerning the description of the action in terms of its results.

I believe it is reasonable to conclude, in light of the previous arguments, that Aristotle’s account of acting through factual error is a unitary account because it is the account of a sort of error which is eminently epistemic, that is, an error involving a positively mistaken factual belief or merely partial factual belief in the senses specified.

B. ACTION-DESCRIPTIONS AND THE PRIMACY CLAIM

Aristotle seems to recognise that whenever I do something through factual error (*di’ agnoian*) and thus involuntarily, I do some other things voluntarily. He notes, for instance, that when someone strikes another person:

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 could have believed truly that the instrument 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,

¹⁴ Rowe’s translation.

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 very plausible to think that this is partly what he means in these passages (see also *MM* 1188b31–6). 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 recognises, 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 offer a love-potion voluntarily, doubtless because doing *x* implies that '*x*' corresponds to a true description of the action. 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 belief.¹⁶ Notice that this does not imply that the woman gave the man a love-potion *involuntarily* either. In fact, '*agnoia*'

¹⁵ *To epistasthai, to eideinai*: *EE* 1225b2, b11; *EN* 1111a23, 1135b20.

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. 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*. If the knowledge-requirement is the requirement that the action be known to the agent under a true description, 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.

(unawareness) is eminently a privative concept, and Aristotle conceives of *agnoia* as *lack* of knowledge (e.g. *ouk eidoûs*, *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 given for her having offered a love-potion, for the drink was not a love-potion. Hence, the right verdict is simply to deny that there was such a thing as offering a love-potion in the first place. As a consequence, voluntariness is description-relative only in the sense that the action-description under which the action turns out to be voluntary corresponds to a *true* description of the action (however partial this could be) which is known to 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 to 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 not only that it is 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 (i.e. result) involuntarily. Aristotle recognises this too (the passage is numbered for ease of reference):

(A) it seems that the person who acts being unaware of *any* (*tî*) 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 *what* and the *how* *heneka*.¹⁸ (*EN* 1111a15–19)

¹⁷ Merely recognising that 'knowingly' and 'not knowingly' are description-relative in the sense specified is not satisfactory unless one has a way to prevent 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 was hemlock, then I offered that hemlock to my father voluntarily. Aristotle seems to have a way to prevent 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 1986: 131–2.

¹⁸ I depart here from the text established by Mingay and Walzer (1991), by inserting '*ho*' before '*kai*' at a18 (see note 20 *infra*). I leave '*hou heneka*' untranslated because I need to justify a certain English translation.

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*)' (IIIb19–22) without any specification as to the relevant description of 'the action' in terms of some circumstance, is perfectly compatible with the fact that what the agent 'did' is involuntary only relative to such description: such unqualified talk can be fully explained by the fact that blame (and 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* IIIa15–19 (*supra*) has been interpreted by J. L. Ackrill 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.¹⁹ 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 IIIa17 ('most precisely' in my translation), which Ackrill seems to take as implying that the agent acts 'most involuntarily' when the error concerns the main conditions (and conversely 'most voluntarily' when he knows these main conditions). Nonetheless, when Aristotle says that the person who acts being unaware of any factual circumstance acts involuntarily '*kai malista en tois kuriôtatois*', he does not mean that he acts *most involuntarily* when he is unaware of the main circumstances. Let me explain.

What Aristotle means, 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,

¹⁹ Ackrill 1978: 98.

involuntarily relative to conditions *c*, *d*, *e*, or *f* but voluntarily relative to condition *a* or *b*.

Passage (B) is about the primacy of two conditions: the *quid* or *what* (*ho*) of the action (its definition) and the *hou heneka* of the action (IIIIa18–19).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 is a possible interpretation of the primacy of *hou heneka* is clear. Aspasius in his commentary argues that all the conditions presuppose unawareness of *hou heneka* in the sense of ‘result’, ‘effect’ or ‘outcome’, as opposed to ‘intention’ or ‘purpose’. Given that the second, and not the former, are the stock translations of ‘*hou heneka*’, Aristotle’s terminology is confusing. It looks as if he thinks that whenever S Φs in order to Z, but instead of Z his Φing results in W, then one can describe the situation as ‘S was unaware that his Φing 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 it will not 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

²⁰ The interpretation of the text has been disputed since ancient times. All the MSS read ‘*en hois hē 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 hē praxis*’ (‘the things in which the action occurs’, EN IIIIa16, 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 hē praxis*’ has been unambiguously used to designate all the conditions identified by Aristotle, including *hou heneka* (IIIIa5; 1135a26), then it is difficult to explain why Aristotle refers to this latter condition as if it was not already covered by ‘*en hois hē praxis*’. It is doubtless because of this that some translators (e.g. Gauthier and Joliff, and Rackham) have adopted H. Richard’s emendation <ho> at a18 and inserted a comma before it (‘*en hois hē praxis*, <ho> kai hou heneka’). Besides, inserting the <ho> seems to be the most economical emendation. This was indeed Nemesius’ (circa 390 AD) interpretation in his *Peri phuseōs anthrōpou* (see Ellebaudt 1565: 129). I think Richard’s emendation and Nemesius’ interpretation are correct. It is worth noting, however, that this emendation was first challenged by Aspasius, who took ‘*en hois hē praxis*’ in its specific sense (cf. *peri ti ē en tini*, IIIIa5), thus considering, along with most commentators, any emendation unnecessary. I find this latter option unviable because just two lines before at a16 ‘*en hois hē praxis*’ is clearly meant in its general sense (and deleting it at a16 as Ramsaurer did is rather unfeasible, again for reasons of economy).

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 actual 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 through factual error is the following:

S was unaware that his *Φing* ____ (e.g. 'with that . . .', 'on that . . .', 'in the way in . . .', etc.) which S mistakenly believed to be X, would result in Z²²

(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 'Φing' 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 Φings (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 *IIII*17–19 is to take it as encompassing two primary conditions, the *quid* of the action – *what* the action is – and its actual, unintended result. We can now account for the primacy of the *quid*. 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*

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 Chapter 7).

²² It is interesting to note that, with only the exception of *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 Quirk *et al.* 1972: ch. 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.

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 listed in *EN* V 8 1135a24–6, *EE* II 9 1225b2–6 and *MM* 1195a17–19. 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 *Ethica Eudemia*, after mentioning the three conditions above and repeating the first two, replaces condition (iii), the *heneka tinos* or actual result, with (iii*) for the *quid* (1225b6–7). Hence, there is some textual evidence to think that Aristotle was inclined to consider these two conditions as identical. And the reason is simple: what 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.

C. CULPABLE ERROR

It is undeniable that Aristotle recognises *culpable* factual error. Indeed, he develops a groundbreaking account of culpable factual error in *EE* II 9, *EN* III 5 and *EN* V 8. Offering an interpretation of this account will be the aim of this and the next sections. 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* 1195a31–4 and *EN* 1113b30–4 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 might 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

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.

information (*anagkaion*, *EE* 1225b15), by which he probably means that the information is such that people are expected to know, both dispositionally and operationally; such information seems to be what the *EN* III 5 characterises 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 (e.g. as part of one’s job, one’s position as a father, or as a citizen) 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 rather 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 a 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 is to be found in *EN* III 1, where he maintains that acting through factual error (*di’agnoian*) is sufficient for rendering the resulting act involuntary:

As factual error is possible with regard to *any* of these conditions, *a person who acts through factual error with regard to any of them is considered as acting involuntarily.* (*EN* 1111a16–18)

This claim has seemed utterly implausible to some: because of it, Aristotle has been accused of being ‘too lenient about errors of fact’ (A. Kenny).²⁴

²⁴ Kenny 1979: 52. See also Williams 1990: 4.

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 a thing), but rather because, even if Aristotle recognised that the culpable agent is to blame for the unawareness, he ought to make this agent distinctly culpable also for the resulting act (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 act, because on Aristotle’s view:

- (I) praise and blame track voluntariness (*EE* 1223a9–15, 1228a10–12; *EN* 1109b30–4; *MM* 1187a20–1), and
- (II) *EN* 1111a16–18, taken at face value, says that all cases of factual error render the resulting act 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 he accepts claim (I), claim (II) (*EN* 1111a16–18) is *implicitly restricted to non-culpable cases*. According to this view (R. 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 in (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’):

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

²⁵ Heinaman 1986: 134 (emphasis added). See also Roberts 1989: 24. Kenny also considers this alternative, but he is hesitant (Kenny 1979: 52–3).

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 blameable (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 either 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 recognise cases of culpable error, and accordingly, he ought not to abandon either 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) *supra*). Finally, I will be suggesting the liberal view about culpable error is not necessarily too lenient.

D. ETHICA EUDEMIA II 9

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

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

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

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 *Ethica Eudemia*? 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 emphasise 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 *ENV* 8 (= *EE* IV 8) shows that actions done *in* error but *not* through error 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

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. 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 two expressions, as employed by *EE* II 9 and *ENV/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

²⁷ Rowe’s translation with slight modifications.

unawareness, for which one can be ultimately 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 emphasises the fact that, when these further causal factors are present, they are dispositions of the agent the salience of which does not call for sympathy, but rather for blame (i.e. he 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 the agent 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 (i) all cases of involuntariness due to factual error, and '*di'agnoian*' to mark (ii) their non-culpability. Conversely, in *EN* III 1 (as we shall see in section E below) Aristotle uses '*di'agnoian*' to mark (i), and seems to have no way of designating (ii).

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 *EE* 1225b11–16: '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 merely 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 claim 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 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.

²⁸ This is important. One strategy available to the conservative interpreter, 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 activated it, knows *in a way* that this person is his father. If one understands b-knowledge as the knowledge that S possesses which is inactive, but which S could easily activate, then it is not 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: 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 unhealthy food *voluntarily* (*EN* 1152a16–17). This is not the sort of b-factual knowledge the conservative interpreter needs, however.

Now, I have argued in this section that according to *EE* II 9 and *EN* V/*EE* IV 8, to say that S acts 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* beyond S's factual unawareness for which S 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 acting 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 the agent's 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 either 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 signalling 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 (IIIIa16–18) 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. *ETHICA NICOMACHEA* III 1 AND 5

EN III 1 (IIIIb24–IIIIa2) also draws a distinction between acting *through* factual error (*di'agnoian*) and acting *in* error (*agnoôn*), examples of the latter being the agent who is drunk or angry. One may think that 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. Right after drawing the distinction (a) between acting through factual error and acting in error, Aristotle goes on to distinguish (b) between ignorance of the universal (*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,

²⁹ As A. Kenny does in Kenny 1979: 49.

the transition from the ‘through factual error’/‘in error’ distinction to the ‘factual unawareness’/‘general unawareness’ distinction is marked by the phrase ‘*agnoei men oun*’ (IIIOb28), 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, but rather because he is ignorant of the universal, normative premise of the practical syllogism; an ignorance that is partly *constitutive* of his faulty character. Think about someone whose drunkenness or anger makes him believe at that time that it is alright to beat his own father.³⁰

Where Aristotle does consider 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*’ (III3b24–6). Aristotle’s apparent endorsement of this practice may suggest that, on his view, the act that results from factual error is not excusable when the error is culpable, either because it is voluntary, or because blameworthy, its involuntariness notwithstanding. 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’ (III3b30–4). 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 (cf. *ep’ autô[i] tô[i] agnoein*, *EN*

³⁰ My claim that the distinction ‘through error/in error’ and the distinction ‘factual ignorance/ ignorance of the universal’ are the same, is challenged by Susan Sauvé Meyer, who claims that ‘the distinction between acting “in ignorance” and acting “because of ignorance”, if Aristotle intends the same distinction in *EN* V 8 and *EN* III 1, does not simply correspond to the distinction between acting with universal and acting with particular ignorance’ (Sauvé Meyer 1993: 178 (emphasis added)). For the relevant passage of *EN* V 8, see section D above. But why suppose that the *same* distinction is intended in *EN* V 8 and *EN* III 1? Earlier in the same chapter, Sauvé Meyer has described *EN* V 8 ‘as a book common to the *EE* that repeats an account of voluntariness similar in many respects to that of *EE* II 9 and appears to be referring back to it’ (*ibid.* 172). I agree with this. Indeed, it is precisely because of this close relationship between *EE* II 9 and *EN* V 8 that there is no reason to suppose that Aristotle intends the same distinction in *EN* V 8 as in *EN* III 1, rather than as in *EE* II 9 (see section D above).

III3b30) 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 *Ethica Eudemia* makes by maintaining that in cases of culpable error, the agent is 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 either 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 used 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: 'Amongst the laws which he made, there 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 These commentaries on the law of Pittacus by Aristotle and Diogenes clearly show that there is no strong evidence that Aristotle thought the penalty was doubled on the ground that there were two offences.

F. CONCLUDING REMARKS

I believe, therefore, that there is no evidence for thinking that Aristotle ever abandons or qualifies the *EN* III1a16–18 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

³¹ For a similar observation, see Charles 1984: 256–8.

³² *Vitae* I 76. Hicks' translation.

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 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 and 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 defendant's acts and the death'.³⁵ The case is structurally similar to the Pittacus example

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 and Ravizza 1998: ch. 7. I think that responsibility as a historical notion is closely connected to responsibility as accountability. Fischer and Ravizza make this clear (*ibid.* 187 and 194–200). The Strawsonian view (which they endorse) 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 and Honoré 1985: 95.

³⁵ *Ibid.* 99 (emphasis added).

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: '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 they 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, liberal notion of culpable error. The enhancement of blameworthiness comes into the picture, as 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

³⁶ Nagel 1993: 29.

³⁷ *Ibid.* 29 (emphasis added).

³⁸ *Ibid.* 29 (emphasis added).

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 denying this. All he is committed to deny is that this sense of blame is not the relevant one for ethical ascription. Aristotle's view on culpable factual error, therefore, is not too lenient, nor is it rendered implausible by our intuitions about cases of moral luck.

CHAPTER 7

The pain condition

A. 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 acting through 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 *Ethica Nicomachea* 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 voluntary 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*.

This suggestion as to how to understand the significance of the pain condition invites a series of criticisms of Aristotle. Most notably, it seems to be evidence that Aristotle's conception of involuntariness, '*to akousion*', is inevitably infected by an unrecognised ambiguity apparently inherent in the Greek term and its cognates, which can sometimes be translated as 'involuntary', and sometimes, as in the present context, as 'unwilling' or '*malgré soi*'.² This sort of criticism is voiced by W. F. R. Hardie, who says

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

² Gauthier and Joliff 1959: vol. II, 169–70.

that this unrecognised ambiguity ‘seems to underlie some of the things which we find perverse or confusing in the text’, citing as an example of such perversity ‘the suggestion that a man whose action is plainly involuntary cannot be said to have acted *akôn* unless he regrets it afterwards (IIIob18–24)’.³ The latter is a reference to the pain condition as applied to involuntariness due to factual error (i.e. subsequent pain or regret), but a comparable uneasiness originates from its application to involuntariness due to violence. We saw in Chapter 3 that the pain condition and the contrariety condition, according to which behaviour brought about through violence is painful and contrary to the organism’s impulse (*hormê*), were constitutive of robust violence. Nonetheless, we also saw that Aristotle is committed to recognise 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. Hence, one might rightly wonder whether Aristotle is once more being misled by this unrecognised ambiguity in the term ‘*akousion*’ and its cognates.

In what follows I should like to resist this line of criticism, by arguing that the best way to understand the robust notion of violence is to see it as conjoining two different conditions into one: the pain condition, affecting the voluntary status of the *agent*, and the externality condition, affecting the voluntary status of the *action*. More generally, I want to argue that the pain condition is relevant *for ethical ascription*, not because it affects the voluntary status of the action, but rather because it directly affects the voluntary status of the agent. I think the same can be said about factual error, so let me start by justifying these claims in relation 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* IIIob19) having revealed the mysteries, and there is the case of someone who does not regret having revealed them at all. In the *Ethica Nicomachea* Aristotle distinguishes these two ‘types’ of acting through factual error:

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

³ Hardie 1980: 152–3.

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;⁴ for since he is different, it is better that he should have a name to himself. (EN IIIIob18–24)

Aristotle's use of '*ouch hekousion*' (lit. 'non-voluntary') and '*akousion*' ('involuntary') in this passage is unprecedented (this is why I have left them untranslated, 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*]',⁵ the latter two being his focus in this passage. On this interpretation, Aristotle is using '*ouch hekousion*' to designate only one of two categories of involuntary 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*' to designate those involuntary actions due to factual error that *are* thus accompanied by pain. But this is not what Aristotle is doing with these two notions. First, he is saying that *every action* that is the result of factual error is *ouch hekousion* (IIIIob18). '*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 I am now proposing, that it is incorrect, because later in the passage Aristotle uses the phrases '*ou peprachen hekôn*' ('has not acted voluntarily') and '*ouch hekôn*' (IIIIo20–I and 23, respectively) to designate the action due to factual error that is not subsequently followed by regret (i.e. the action done 'non-voluntarily').⁶ 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 species (i.e. what is due to factual error and not subsequently accompanied by pain) – but this is surely implausible.

⁴ I follow here most commentators in taking the negation '*ouch*' with '*hekôn*' rather than with '*estô*' (the same in a18).

⁵ Sauvé Meyer 1988: 129. Her interpretation is partly supported by her taking the *ouch* at a18 with '*estin*' ('is not voluntary', *ibid.* 127) rather than with '*hekousion*' ('is non-voluntary', as Rowe translates it). The latter is much more plausible given the position of the words.

⁶ Sauvé Meyer, for instance, reads '*ouch hekôn*' at a23 adverbially ('[he acts] non-voluntarily', *ibid.* 128), and the same with '*akôn*' at a21.

In spite of these appearances, this is not what Aristotle is doing with these phrases. First, Aristotle is using ‘*ouch hekôn*’, as distinct from ‘*ouch hekousion*’, to designate, not the action, but the agent: the ‘non-voluntary agent’ who does not feel subsequent regret for having acted through factual error, when learning the relevant facts (cf. *ho mê metamelomenos*, b23). Likewise, he is using ‘*akôn*’, as distinct from ‘*akousion*’, to designate the agent who feels subsequent regret for having acted through factual error (cf. *ho en metameleia*[i], b22).⁷ 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 characterise the agent on the basis of his attitude towards that involuntary action. He can do this because the Greek adjectives ‘*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 seems now to be using ‘*akôn*’ and ‘*ouch hekôn*’ (as adverbs) to qualify actions. But notice his careful use of the perfect tense.⁸ 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 assumption to make.⁹ For whereas it is true that ‘S has Φ ed’ can be used merely to report past experiences (as in ‘S has been to Chile a couple of times’), thus referring to past occasions of Φ 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

⁷ Ross’ translation captures this nicely.

⁸ On Aristotle’s use of the perfect in this connection see also Broadie 1991: 139–40.

⁹ Potts and Taylor 1965: 68.

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’.¹⁰ Thus understood, sentences of the form ‘S has Φ ed’ are sentences that can be true only in virtue of present conditions. In the case of ‘S has Φ ed *akôn/ouch hekôn*’ these present conditions include the agent’s present attitude towards his past action. The use of the perfect then indicates 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 actions in (i). 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 easily be extended¹¹ so as to cover all agents who feel the relevant sort of pain for acting or 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 category of the ‘*ouch hekôn*-agent’ (from now on, ‘the non-voluntary agent’) should be extended to cover all those agents who do not feel the relevant sort of pain for acting or having acted involuntarily. No doubt, the legitimacy of this extension can be disputed by the fact that such a 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

¹⁰ Potts and Taylor 1965: 86–7.

¹¹ As Hursthouse 1984 and Irwin 1988 do.

something which is not contrary to your internal impulses, so there is no room for non-voluntary action due to force'.¹² I have argued in [Chapter 3](#), section F, that this is not so: Aristotle recognises, and is committed to recognise, a core definition of violence. Because of this, he is plainly committed to acknowledge a distinction analogous to the one between the two 'sub-categories 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 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 *x* that is performed involuntarily, an agent *S* is now *involuntary* (*akôn*) with regard to *x* if *S* now feels pain for causing or having caused *x*; and *S* is now *non-voluntary* (*ouch hekôn*) with regard to *x* if he now feels no pain for causing or having caused *x*.

Now, when I suggested at the beginning of this section 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 error', tell us whether the piece of behaviour *x* is involuntary in the first place; that is, the significance of the core definition of violence and the 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 vis-à-vis *x*. The robust definition of violence misleadingly conjoins (i) and (ii).

The precise meaning of this principle, PC, will not become clear until section D below. Next, we must turn to an account of the sort of pain that characterises the involuntary agent, according to Aristotle, but before that we must polish our account of the pain condition.

B. REFINING OUR ACCOUNT OF THE PAIN CONDITION

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 a contrariety condition, i.e. 'against *S*'s impulse (*hormê*)'. If the relation between the

¹² Whiting 1989: 163–4. See also Means 1927: 83–4.

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), 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. 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.'¹³

It is, however, very significant that Aristotle's 'robust' conception of factual error only involves the pain condition. For all Aristotle says in *EN* 1110b18–24 and anywhere else, it is *the presence of regret* that makes an agent S involuntary (*akôn*) vis-à-vis a piece of behaviour x caused by his factual error, not the fact that x has been contrary to S's impulse all along. What makes x itself involuntary in cases of factual error, in turn, is merely the fact that x is the product of factual error. 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, but it does imply, I think, that to the extent that there is some contrariety to impulse in cases of factual error, such contrariety and the corresponding pain stand to each other at the very least in a relation of mutual implication. Perhaps one could make the stronger claim that pain (adequately understood) is just what 'contrariety to impulse' *means*.

My interpretation of the pain condition is based on the characteristically moral pleasure and pain that according to Aristotle are part of the exercise of virtue itself, or of vice: 'of the greatest importance for virtue of character is to find pleasure in the things one must and to hate the things one must' (*EN* 1172a22–3). I discuss these distinctively moral pains and pleasures in the next section, but the point I wish to make here is that these moral pains and pleasures are *constitutive* of the fact that the action targeted by

¹³ Sauvé Meyer 1988: 133.

them, is itself *according to* or *contrary to* the agent's moral quality. If this is the sort of pain that makes an agent involuntary vis-à-vis an ethically significant piece of behaviour *x*, there is then no conceptual space for the assertion that *x* is according to, or contrary to the involuntary agent's moral quality, independently (conceptually speaking) of the agent's being pleased or pained at *x*.

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 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 *x*, if (i) S feels the relevant pain for causing *x*, or (ii) S would feel the relevant pain if S were to know the relevant facts; and S is a *non-voluntary agent* towards *x* 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 hypnotised 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-hypnotised 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 hypnotised 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* 1109b30–1), where ‘*epi tois hekousiois*’ seems to refer back to both affections (*pathê*) and actions (*praxeis*). But what does Aristotle mean by ‘affections’ (*pathê*) here? I will try to answer this question in the present section, and in section E below I shall suggest a sense in which these affections can be ‘voluntary’.

To get hold of the relevant sense of ‘affection’ we need to examine the claim that ‘excellence has to do with affections 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; *EE* 1220a34). One of its most natural interpretations is that each virtue (perhaps with the sole exception of justice) has a specific domain of emotions within which it is expected to operate, by disposing the agent to be affected by them in the appropriate ways. For our present purposes, however, there is a more significant sense in which affections figure amongst 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).¹⁴ Pleasures and pains, in so far as they are the concern of ethical excellence (i.e. ‘moral’ pleasures and pains), are clearly amongst 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), as much 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)¹⁵ the particular ways of being affected and acting that are peculiar to a given ethical disposition. Therefore, our initial claim that ethical dispositions are concerned with

¹⁴ Also *EE* 1220a34–9, 1221b35–9, 1227b1–2, 1227b5–6; *MM* 1185b37.

¹⁵ Also *EE* 1220b14; *Rhet.* 1378a21; *MM* 1186a13. The verb ‘*hepomaî*’ 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.

affections (*pathê*), can be narrowed down to the claim that each type of ethical disposition is concerned with *the pain and pleasure that supervene on*, or follow upon, the emotions and actions characteristic of it – its operations (*erga*). What is vital for my present purposes is that these pleasures and pains not only accompany the emotions characteristic of each ethical disposition (every emotion *is a kind of* pleasure or pain, *EE* 1221b37; *Rhet.* 1378a21), but also accompany actions, *praxeis* (*EN* 1104b14). This extension of moral pains and pleasures which are the concern of virtue, to the domain of action, will be the focus of attention in what follows.

Aristotle has already noted in Book I of the *Ethica Nicomachea* 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 he has 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 noting Aristotle’s explicit recognition 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 recognises 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 (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 *EN* 1104b5–8 passage just quoted. 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 describe 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

¹⁶ Rowe’s translation. Aristotle 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* 1101b11–13).

¹⁷ *EE* 1224a31, a37, b17, b19, 1225a25, 1227b10.

1104b5–8 *supra* is taken to be the one between the continent (rather than the self-indulgent) and the moderate person, surely the contrast between the cowardly and the brave in the same passage is between (i) someone who does withstand frightening things but does so reluctantly (as opposed to the perfect coward who does not withstand frightening things because the prospect of doing so is too painful), 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), Aristotle's point here is that 'self-indulgent' and 'cowardly' in this context are the contradictories of 'moderate' and 'courageous', respectively, 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 us with a model to understand what Aristotle means by 'involuntary agent' and 'non-voluntary agent' in PC*.

Now, some important 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 involuntary (*akontî*)' (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 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 ought not to have the unpalatable result that he is also pleased or undisturbed by getting wounded

¹⁸ Rowe's translation.

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) These are the sort of attitudes that are directed at the ethically salient aspects of an action or event.¹⁹

This feature of moral aversion- and compliance-attitudes implies that these 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 view the pain and lack of pain that make an agent involuntary and non-voluntary vis-à-vis his involuntary action, also as moral aversion- and compliance-attitudes, respectively. This view is possible because involuntary actions (or motions) *can also have ethically salient aspects*, according to Aristotle. Aristotle recognises, for instance, that an involuntary action or state of affairs can be described as 'something unjust' (*ti adikón*) or as 'something moderate' (*ti sophrón*), while recognising that it is not thereby blameworthy or praiseworthy (*EN* 1135a9–11, a19–23, 1105b5–10).

Interestingly enough, there is another feature of moral aversion- or compliance-attitudes (now restricted to 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 might not only feel pain against striking C, but also against being forced by A, notwithstanding his attitude towards the striking of C.²⁰ So another feature of the moral aversion- and compliance-attitudes directed at involuntary behaviour is this:

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/vice-terms.

²⁰ 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 not to notice that pain can directly relate to the externality condition

(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 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 it 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 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.

Let us consider once more the case of a diabetic person who craves 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. It is now evident that the rational pleasure he experiences 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 pain felt by the diabetic person exhibits his divided condition, his continent or incontinent condition. It is clear that the relevant attitude, for the purposes

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).

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 wellbeing, rather than his quasi-biological pain pointing in 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) *supra*), and given that it is implausible to think that biologically-based cravings and non-rational impulses can discharge this function, it is reasonable to think that the impulse relevant to the robust notions of violence and factual error is the agent's rational impulse (*boulêsis*) aiming at the good (*Rhet.* 1369a2–7), as suggested in Chapter 3, section F.²¹ 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 such a domain either; as a matter of fact, 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 (cf. *EE* 1222a6–14, 1227b5–10). Even though he does not say so explicitly, Aristotle would be justified in claiming that only agents capable of holding compliance- and aversion-attitudes (thus defined) can be *voluntary agents*. Furthermore, it is natural to think that he would identify these agents with *reason-responsive agents*, since any reason-responsive agent can in principle exercise the discriminatory capacities required by the holding of these moral attitudes.

D. 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). One is certainly entitled to think that there is a close connection between someone's motivation and, say, the moral aversion-attitude he takes towards an event brought about in his body through violence. It is tempting to think that

²¹ 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.

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 prominence of the motivational aspect of aversion-attitudes that such attitudes 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.²² But is this really true after all?

I don't think so. 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, 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 life is rather comforting. But this should not prevent us from concluding that, if two weeks after the incident the father helped another criminal escape involuntarily believing falsely that he was a tourist asking for a lift, and subsequently regretted the outcome upon learning the truth, such 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 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 to say, it 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

²² This interpretation is not uncommon. See, for instance, Moline 1989.

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 experiences *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 characterises 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 deliberation, 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 whenever the agent is confronted with a piece of behaviour of his that is at variance with this evaluative conception, whether such 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.

In addition to these considerations, notice 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 in 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. Suppose 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 A is rather glad of 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 A's lack of distress is the relevant sort of pain, that is, if it corresponds indeed to a moral compliance-attitude, then it tells us something about A's evaluative outlook on B's accident.²³

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*:

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, that is, we usually assume [1]. We see the cavalryman fleeing from the battlefield, and assume this to be a voluntary, cowardly action. Notice that the cases where 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 performed the action in full knowledge of the 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' and thus to move on to step

²³ It seems to me that one can see the significance of this identification or disassociation attached to moral compliance- and aversion-attitudes 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 of them, satisfaction at their being done, or perhaps simply by being pained at their being done, just as one can dissociate oneself 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.

[2]? On Aristotle's view, we are not, unless we have also established that he 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 battlefield because he had to attend to a game of poker. Although this may be a suitable basis for calling him 'irresponsible', it is not a basis for blaming him as '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 behaviour these attitudes are directed at is involuntary (or perhaps even caused by some other agent), and even though they do not necessarily tell us what the agent would voluntarily do in a counterfactual scenario. According to my interpretation, aversion- and compliance-attitudes can be attitudes towards involuntary pieces of behaviour all the same, because even involuntary pieces of behaviour can have ethically salient features. We have seen in the previous section that Aristotle acknowledges this.

I am now in a position to maintain that, if a putative agent has a moral aversion-attitude towards an involuntary piece of behaviour *x* and its ethically salient features, e.g. as something unjust (*ti adikón*), this of course defeats the assumption of voluntariness at level [1] in the most successful way possible, namely, by signalling the fact that he is a *non-unjust* agent: this is the force of 'involuntary agent'. A compliant moral attitude towards *x* as something unjust (*ti adikón*), 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 an involuntary piece of behaviour are good (though not 'praisable'), e.g. something moderate (*ti sôphron*). Here, a moral aversion-attitude towards *x* tells us something more positive about the agent: although we cannot call him 'immoderate', we know he is a *non-moderate* agent. If his attitude is a moral attitude of compliance, on the other hand, although we cannot 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 *x*, whereas

compliance attitudes are compatible, as our interpretation of *EN* 1104b5–8 in section C above suggested.²⁴

Accordingly, there is a precise way in which moral compliance and aversion-attitudes are ‘voluntary’ *pathê* (cf. *EN* 1109b30–1), namely, they directly affect the voluntary status of the agent, without the mediation (*dia*) of the agent’s voluntary behaviour. I take it that doubts that may arise as to why exactly the cavalryman’s reluctance and distress at his being dragged away from the battlefield should be incompatible with his being a coward, are the more acute the more unspecified the reluctance and distress in question remain. We have seen two ways in which this can be so, corresponding to the two features of moral attitudes distinguished in section C above: (i) they 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 battlefield*, 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 battlefield in so far as this is something bad and cowardly, or in so far as the experience at the battlefield involves things the cavalryman 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, therefore, turns out to be: ‘why should we think that the cavalryman’s reluctance and distress at his being dragged away *from the battlefield, in so far as* leaving the battlefield on this occasion is seen by him *as something cowardly*, is incompatible with his being a cowardly person?’. Thus expressed, the answer is virtually contained in the question.

²⁴ A diagram would look like this:

Assumed character → Moral attitudes ↓	Virtuous character (on the basis of e.g. <i>ti dikaion</i>)	Faulty character (on the basis of e.g. <i>ti adikōn</i>)
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)

Conclusion

In Part I I have established that Aristotle's discussion of voluntariness in his *Ethics* is grounded on a viable conception of moral responsibility, which I have termed 'ethical ascription'. The praise and blame involved in ascribing a voluntary, ethically significant action to an agent are neither sanction-oriented, backward-looking reactions focused on what is deserved, nor merely instrumental devices for affecting future behaviour; the praise and blame involved concern themselves with what can be ascribed to reason-responsive agents (who *are* moral agents), and have a prospective dimension that is appropriate, as purely instrumental approaches are not, to such beings. Furthermore, I have also established that the ascription of an ethically significant action to a reason-responsive agent is carried out from the normative stance of the virtuous spectator (Chapter 2, section D).

Now, on this basis we can understand Aristotle's detailed account of the voluntary and the involuntary as an explanation of the conditions that defeat the ascription of ethically significant behaviour to such reason-responsive agents. Examining such an explanation has been the focus of Part II. This examination has revealed the intricate and complex nature of Aristotle's accounts and brought to the surface a great variety of problems that have had to be solved in order to arrive at a satisfactory interpretation. Nonetheless, once a satisfactory interpretation of Aristotle's accounts of the various defeaters of ethical ascription has been achieved, we can see that such accounts are much richer and more complex than is commonly supposed. Readers of Aristotle's *Ethica Nicomachea* are usually satisfied with its definition of involuntariness as 'what comes about through violence and what comes about through unawareness' (IIIIa22), while failing to note that he developed a position concerning not only violence, but also coercion, and not only coercion, but coercion as justification and excuse; or that he also has a view, not only of factual unawareness, but of culpable factual unawareness. Surely, Aristotle's definition of involuntariness is formulated

in such simple terms because he is eager to derive from it the definition of voluntary action, as that which has an internal origin within the agent when he knows the particular, relevant facts surrounding his action; but as I have stated in the Introduction, and as the reader can now infer from the discussion in [Chapter 3](#), this cannot be done.

Once we arrive at a satisfactory interpretation of Aristotle's accounts of the various factors that defeat ethical ascription, we can also see that these are, in the end, *coherent*. Readers who are not satisfied with the definition of involuntariness in *EN* IIIIa22 are usually troubled by the apparently new definition of 'involuntary' in *EN* IIIIb18–24, as involving *pain*. Moreover, if they enquire further, they perceive a difficulty in Aristotle's requirement in the *Ethica Eudemia* that an involuntary action due to violence requires *contrariety to impulse*. Indeed, not only is pain not included in the definition of involuntary action in *EN* IIIIa22, but it now appears as though Aristotle is not concerned with the concept of involuntariness after all, but rather with the concept of *unwillingness*. The distinction which I have made in [Chapter 7](#) between involuntary action and the involuntary agent reveals the former worry to be unfounded. My argument in [Chapters 3](#) and [7](#) that the Eudemean 'robust' notion of violence (misleadingly) conjoins these two notions shows the second worry to be unwarranted. Other minor objections relating to incoherence are likely to be levelled at different junctures of Aristotle's discussion, such as the apparent inconsistencies in his view of culpable unawareness; but I have shown Aristotle's texts to be consistent in this regard.

Readers are also likely to be troubled, if not by the incoherence of Aristotle's discussion of involuntariness, at least by the disagreement between the Eudemean and the Nicomachean discussions. Nonetheless, once we see that in the respects in which these two treatises evidently disagree the Nicomachean position is superior, such disagreements should no longer present a difficulty; for now we have a normative criterion to resolve them. At least two examples of the Nicomachean superiority can be gathered from my discussion: where the Eudemean discussion of coercion fails to do justice to its justificatory role ([Chapter 4](#)), the Nicomachean acknowledges it ([Chapter 5](#)); and where the Eudemean discussion misleadingly conjoins the involuntariness of the action and the involuntariness of the agent, the Nicomachean succeeds in distinguishing them.

I also believe Aristotle's account to be interesting and illuminating in many regards. His notion of moral responsibility as ethical ascription, for instance, reveals some of the shortcomings of Watson's and Feinberg's conception of attributability, such as the unjustified assumption that

attributability appraisals are to be held privately. Furthermore, Aristotle's own conception addresses, in a way that is original and challenging, questions that their conception of attributability does not explicitly address, such as the question concerning the sort of authoritative stance that ought to be adopted in attributing an ethically significant action to an agent, and the question concerning the sort of agent that can be the adequate recipient of such attribution. Aristotle's account of those factors that defeat the ascription of ethically significant behaviour to reason-responsive agents is also interesting and illuminating in many respects. His discussion of coercion seems to account in original ways for some of the things we find puzzling about coercion, such as the ambivalence between its subjective and objective nature, and between its role as both an excuse and a justification; his account of coercion as *excuse* is indeed superior, I believe, to those that locate its exculpatory force in the peculiar psychological make-up of an individual. His discussion of factual ignorance is plagued with both implicit and explicit distinctions, usually unacknowledged, that should be of the greatest interest to any theory of excuses and contemporary theories of action, such as the distinction between positively mistaken and merely partial belief, or the distinction between those facts regarding an action that are primary and those that are secondary. His view of culpable ignorance is firmly grounded, I believe, in his peculiar conception of moral responsibility as attributability. It remains to be seen whether such a conception can find its way into the contemporary debate on moral responsibility. To quote the final words of the *Ethica Nicomachea*, 'Let us then make a beginning of the discussion.'

*Two lexical ambiguities in Aristotle's use of
'to eph' hautô[i]'*

In this appendix I attempt to clarify two lexical ambiguities involved in the notion of *to eph' hautô[i]*, 'what depends on one' or 'what is in one's control', 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)

But (1) and (2) are unsatisfactory formulations of the notion of *to 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 dependent, are indeed different. This is misleading because Aristotle clearly believed that *the same* contingent power enabled one both to Φ and not to Φ . He thought that if I indeed possess at *t* the contingent 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 contingent power.² Formulation (2), on the other hand, suggests that there is only one contingent power at stake. The problem with (2) is that it is misleading because it also suggests that there is one single contingent power that can be actualised 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*', 'and' (e.g. *EN* 1110a16). In what follows, I will use the disjunction '*ē*', 'or' (e.g. *EE* 1223a7–8):

(3) Φ depends on S = It depends on S (to Φ V not to Φ).

I want to argue that (3) is indeed the correct formulation of Aristotle's notion of *to ep' hautô[i]*.

¹ E.g. *bosa d' eph' hautô[i] esti poiein ē mê poiein* (*EE* 1223a7–8); *ep' autô[i] kai to prattein kai mê* (*EN* 1110a16); *en bois 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.

II. In order to show this, still another ambiguity must be clarified: there is another sense in which the notion of *to 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 accordance with (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 *to eph' hautô[i]* in connection with the Platonic question about the voluntariness of virtue and vice, he uses a formula of the (4)-type *supra*. But in fact, Aristotle never says this explicitly. On the contrary, I want to show that in these contexts, it is a notion of *to eph' hautô[i]* captured by (3), and not by (4), that Aristotle uses and needs.

EN III 5 argues as follows (sentences are numbered for ease 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 assuming a notion of *to eph' hautô[i]* as captured by (3) *supra*. To show this, let me first rule out one possible interpretation

³ Rowe's translation.

of the passage. It is tempting to think that Aristotle is assuming in the *EN* 1113b6–14 passage a principle such as the following:

P: It depends on us to $\Phi \leftrightarrow$ 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 (cf. *EN* 1113b13), as Plato 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 (call it ‘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 $\Phi \rightarrow$ It depends on S *not* to Ψ

2 It depends on S *not* to $\Phi \rightarrow$ It depends on S to Ψ

3 *Ergo*: It depends on S to $\Phi \rightarrow$ 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) *supra*, and a simple argument to show that the Asymmetry Thesis is false:

1 It depends on S to $\Phi \rightarrow$ It depends on S *not* to Φ (3)

2 It depends on S *not* to $\Phi \rightarrow$ It depends on S to Ψ

3 *Ergo*: It depends on S to $\Phi \rightarrow$ It depends on S to Ψ

⁴ Salomon’s translation. ⁵ See Plato’s *Protagoras* 345d–e; *Timaeus* 87b; *Laws* 86od.

⁶ The *locus classicus* is *Laws* IX, 86od. On this see Sauvé Meyer 1993: 131–2.

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 *unjust*, precisely by not returning the deposit. I conclude, then, that all Aristotle needs as a definition of *eph' hautô*[1] is formulation (3).

Bibliography

- Ackrill, J. L. 1978. 'Aristotle on action', in Rorty (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 (ed.) 1980, pp. 285–300
1992. 'Ancient ethics and modern morality', *Philosophical Perspectives* 6: 119–36
- Anscombe, G. E. M. 1958. 'Modern moral philosophy', in A. P. Martinich and D. Sosa (eds.), *Analytic Philosophy: An Anthology* (Wiley-Blackwell, 2001), pp. 381–92. (Reprinted from *Philosophy* 33: 1–19)
1963. 'Two kinds of error in action', *Journal of Philosophy* 60: 393–401
- Araujo, M. and Marias, J. 1970. *Aristóteles: Ética a Nicómaco*. Madrid: Instituto de Estudios Públicos
- Austin, J. 1873. 'Negligence, heedlessness, and rashness', in Morris (ed.) 1961, pp. 240–6. (Reprinted from his *Lectures on Jurisprudence* (London: John Murray), pp. 440–4)
- Austin, J. L. 1956–57. 'A plea for excuses', in Morris (ed.) 1961, pp. 6–19. (Reprinted from *Proceedings of the Aristotelian Society* 57: 1–30)
- Baier, K. 1972. 'Types of responsibility', in French (ed.) 1991, pp. 117–22. (Reprinted from P. French (ed.), *Individual and Collective Responsibility* (Cambridge, MA: Schenkman Publishing), pp. 56–61)
- Baker, M. B. 1983. 'The excuse of accident', *Ethics* 93: 695–708
- Barnes, J., Schofield, M. and Sorabji, R. (eds.) 1977. *Articles on Aristotle*, Vol. II. London: Duckworth
- Bennett, J. 1980. 'Accountability', in Z. Van Ztraaten (ed.), *Philosophical Subjects: Essays presented to P. F. Strawson* (Oxford: Clarendon Press, 1980), pp. 14–47
- Bobzien, S. forthcoming. 'Choice and responsibility in *Nicomachean Ethics* iii 1–5', in R. Polansky (ed.), *The Cambridge Companion to Aristotle's Nicomachean Ethics* (Cambridge: Cambridge University 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 (ed.) 1991, pp. 90–109. (Reprinted from A. I. Melden (ed.), *Essays in Moral Philosophy* (Seattle: University of Washington Press), pp. 3–39)
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: 109–32
- Brickhouse, T. C. 1991. 'Roberts on responsibility for action and character in the NE', *Ancient Philosophy* 11: 137–48
2003. 'Does Aristotle have a consistent account of vice?', *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: 229–42
- Broadie, S. and Rowe, C. 2002. *Aristotle: Nicomachean Ethics*. Oxford, New York: Oxford University Press
- Burnet, J. 1900. *The Ethics of Aristotle*. London: Methuen
- 1905–13. *Platonis Opera*, Vol. 1–5. Oxford: Oxford University Press
- Burnyeat, M. F. 1980. 'Aristotle on learning to be good', in Rorty (ed.) 1980, pp. 69–92
- Bywater, I. 1894. *Aristotelis Ethica Nicomachea*. Oxford: Clarendon Press
- Cairns, D. L. 1993. *Aidôs: The Psychology and Ethics of Honour and Shame in Ancient Greek Literature*. Oxford, New York: Oxford University Press
- Charles, D. 1984. *Aristotle's Philosophy of Action*. Ithaca: Cornell University Press
2009. 'Nicomachean Ethics VII. 3: varieties of *akrasia*', in C. Natali (ed.), *Aristotle: Nicomachean Ethics, Book VII Symposium Aristotelicum* (Oxford, New York: Oxford University Press, 2009), pp. 41–71
- Cohen, S. M. 1994. 'Aristotle on elemental motion', *Phronesis* 39: 150–9
- Cooper, J. M. 1973. 'The *Magna Moralia* and Aristotle's moral philosophy', *American Journal of Philology* 94: 327–49
1975. *Reason and Human Good in Aristotle*. Cambridge, MA: Harvard University Press
- Curren, R. 1989. 'The contribution of Nicomachean Ethics iii 5 to Aristotle's theory of responsibility', *History of Philosophy Quarterly* 6: 261–77
- Darwall, S. 2006. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA, London: Harvard University Press
- Davidson, D. 2001. *Essays on Actions and Events*, 2nd edn. Oxford: Clarendon Press
- Déclarie, V. and Houdé-Sauvé, R. 1978. *Aristotle: Éthique à Eudème*. Paris: Librairie Philosophique J. VRIN
- Destrée, P. 2011. 'Aristotle on responsibility for one's character', in M. Palaluk and G. Pearson (eds.), *Moral Psychology and Human Action in Aristotle* (Oxford: Oxford University Press, 2011), pp. 285–319

- Di Muzio, G. 2000. 'Aristotle on improving one's character', *Phronesis* 45: 205–19
- Dodds, E. R. 1951. *The Greeks and the Irrational*. Berkeley, Los Angeles: University of California Press
- Dover, K. J. 1994. *Greek Popular Morality in the Time of Plato and Aristotle*. Indianapolis: Hackett
- Duff, R. A. 1990. *Intention, Agency and 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: 250–2
- Ellebaudt, N. (ed.) 1565. *Nemesii Episcopi et De Natura Hominis, Lib. Unus*. Antwerp: Ex Officina Christophori Plantini
- Everson, S. 1990. 'Aristotle's compatibilism in the *Nicomachean Ethics*', *Ancient Philosophy* 10: 81–99
- Feinberg, J. 1970. *Doing and Deserving*. Princeton: Princeton University Press
1986. *Harm to Self*. New York, Oxford: Oxford University Press
- Fischer, J. M. and 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: 329–55
- 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 T. Honderich (ed.), *Essays on Freedom of Action* (London: Routledge & Kegan Paul, 1973), pp. 65–86
- Frede, D. 1986. 'Mixed feelings in Aristotle's rhetoric', in Rorty (ed.) 1996, pp. 258–85
2006. 'Pleasure and pain in Aristotle's ethics', in Kraut (ed.) 2006, pp. 255–75
- Frede, M. 2011. *A Free Will: Origins of the Notion in Ancient Thought*, edited by A. A. Long. Berkeley: University of California Press
- Freeland, C. 1986. 'Aristotle on possibilities and capacities', *Ancient Philosophy* 6: 69–89
- Freese, J. H. 2000. *Aristotle, the Art of Rhetoric*. Cambridge, MA: Harvard University Press
- French, P. (ed.) 1991. *The Spectrum of Responsibility*. New York: St Martin's Press
- Furley, D. J. 1967. *Two Studies in the Greek Atomists*. Princeton: Princeton University Press
1977. 'Aristotle on the voluntary', in Barnes, Schofield and Sorabji (eds.) 1977, pp. 47–60
1978. 'Self-movers', in Rorty (ed.) 1980, pp. 55–68. (Reprinted from G. E. R. Lloyd and G. E. L. Owen (eds.), *Aristotle on Mind and the Senses* (Cambridge: Cambridge University Press), pp. 165–79)
- Gagarin, M. 1981. *Drakon and Early Athenian Homicide Law*. New Haven, London: Yale University Press

- Gauthier, R. A. and Joliff, J. Y. 1959. *Aristote: l'Ethique à Nicomaque*, 2nd edn, 2 vols. Leuven: Publications Universitaires
- Gilby, T. 2006. *Thomas Aquinas, Summa Theologiae*, Vol. XVII, *Psychology of Human Acts*. New York: Cambridge University Press
- Gill, M. L. 1991. 'Aristotle on self-motion', in L. Judson (ed.), *Aristotle's Physics: A Collection of Essays*. New York, Oxford: Clarendon Press, 1991), pp. 243–65
- Grant, A. 1885. *The Ethics of Aristotle*, 4th edn, 2 vols. London: Longman
- Grönroos, G. 2007. 'Listening to reason in Aristotle's moral psychology', *Oxford Studies on Ancient Philosophy* 32: 251–72
- Haksar, V. 1964. 'Aristotle and the punishment of psychopaths', *Philosophy* 39: 323–40
- 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 edn. 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 (ed.) 1961, pp. 143–57. (Reprinted from *Proceedings of the Aristotelian Society* 59: 171–19)
- Hart, H. L. A. and Honoré, T. 1985. *Causation in the Law*, 2nd edn. Oxford, New York: Oxford University Press
- Heinaman, R. 1986. 'The Eudemian Ethics on knowledge and voluntary action', *Phronesis* 31: 128–47
1988. 'Compulsion and voluntary action in the Eudemian Ethics', *Noûs* 22: 253–81
- Heybut, G. (ed.) 1889. *Aspasius in Ethica Nicomachea Quae Supersunt Commentaria*, Vol. I. Berlin: George Reimer
- Hicks, R. D. 1925. *Diogenes Laertius: Lives of Eminent Philosophers*, Vol. I. London: Heinemann
- 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 (ed.) 1980, pp. 117–56
1988. *Aristotle's First Principles*. Oxford: Clarendon Press
1996. 'Ethics in the *Rhetoric* and in the *Ethics*', in Rorty (ed.) 1996, pp. 142–74
2001. 'Vice and reason', *Journal of Ethics* 5: 73–97
2007. *The Development of Ethics*, Vol. I. 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: 159–60
- 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 (ed.) 1980, pp. 103–16
- 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 *Nicomachean Ethics*', *Classical Quarterly* 31: 129–40
- Leighton, S. R. 1996. 'Aristotle on the emotions', in Rorty (ed.) 1996, pp. 206–37
- Loomis, W. T. 1972. 'The nature of premeditation in Athenian homicide law', *Journal of Hellenic Studies* 92: 86–95
- MacDowell, D. M. 1966. *Athenian Homicide Law*, 2nd edn. Manchester: Manchester University Press
1991. *Gorgias: Encomium of Helen*, 2nd edn. Bristol: Bristol Classical Press
- Mackenzie, M. M. 1981. *Plato on Punishment*. Berkeley, Los Angeles, London: University of California 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: 139–56
1992. *Springs of Action: Understanding Intentional Behaviour*. New York, Oxford: Oxford University Press
- Mingay, J. M. and 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: Stanford University Press
- Murray, M. J. and Drudrick, D. F. 1995. 'Are coerced acts free?', *American Philosophical Quarterly* 32: 109–23
- Nagel, T. 1993. 'Moral luck', in Statman (ed.) 1993, pp. 57–72
- Natali, C. 2004. *L'action efficace: Études sur la philosophie de l'action d'Aristote*. Leuven: Éditions Peeters
- Nauck, A. 1926. *Tragicorum Graecorum Fragmenta*, 2nd edn. Leipzig: Teubner

- Nielsen, K. M. 2007. 'Dirtying Aristotle's hands? Aristotle's analysis of "mixed acts" in the *Nicomachean Ethics* III, 1', *Phronesis* 52: 270–300
- Nozick, R. 1969. 'Coercion', in S. Morgenbesser, W. Morton and S. Patrick (eds.), *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel* (New York: St Martin's Press), pp. 440–72
- Nussbaum, M. C. 1978. *Aristotle's De Motu Animalium*. Princeton: Princeton University Press
1986. *The Fragility of Goodness*. Cambridge: Cambridge University Press
- Oshana, 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 I. Düring and G. E. L. Owen (eds.), *Aristotle and Plato in the Mid-Fourth Century* (Gothenburg: Studia Graeca et Latina Gothoburgensia XI, 1960), pp. 163–90
- Pakaluk, M. 2005. *Aristotle's Nicomachean Ethics: An Introduction*. Cambridge: Cambridge University Press
- Potts, T. C. and Taylor, C. C. 1965. 'States, activities and performances', *Proceedings of the Aristotelian Society* (Supplement) 39: 65–102
- Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. 1972. *A Grammar of Contemporary English*. London: Longman
- Rackham, H. 1938. *Aristotle: The Athenian Constitution, The Eudemian Ethics, On Virtues and Vices*. Cambridge, MA: Harvard University Press
- Richards, A. and Stratchey, J. 1991. *Sigmund Freud: New Introductory Lectures on Psychoanalysis*, Vol. II. London, New York: Penguin
- Roberts, J. 1989. 'Aristotle on responsibility for action and character', *Ancient Philosophy* 9: 23–36
- Roemer, A. 1922. *Aristotelis 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', *Meetings of the Aristotelian Society* 103: 61–84
- 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 Analytica Priora et Posteriora*. Oxford: Clarendon Press
1979. *Aristotelis Topica et Sophistici Elenchi*. Oxford: Clarendon Press
- Russel, P. 1995. *Freedom and Moral Sentiment: Hume's Way of Naturalizing Responsibility*. New York: Oxford University Press
- Ryle, G. 1949. *The Concept of Mind*. Harmondsworth: Penguin

- Sakezles, P. 2007. 'Aristotle and Chrysippus on the psychology of human action: criteria for responsibility', *British Journal for the History of Philosophy* 15: 225–52
- Sauvé Meyer, S. 1988. 'Why involuntary actions are painful', *Southern Journal of Philosophy* (Supplement) 27: 127–58
1993. *Aristotle on Moral Responsibility: Character and Cause*. Oxford: Basil Blackwell
1998. 'Moral responsibility: Aristotle and after', in S. Everson (ed.), *Companions to Ancient Thought*, Vol. IV, *Ethics* (Cambridge: Cambridge University Press, 1998), pp. 221–40
- 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 (ed.) 2003, pp. 58–71. (Reprinted from *Mind* 70: 291–306)
- Smith, H. 1983. 'Culpable ignorance', *Philosophical Review* 92: 543–71
- Solomon, J. 1915. *Ethica Eudaemia*, in D. Ross (ed.), *The Works of Aristotle Translated into English*, Vol. XIX (Oxford: Oxford University Press, 1915)
- Sorabji, R. 1973–74. 'Aristotle on the role of intellect in virtue', in Rorty (ed.) 1980, pp. 201–20. (Reprinted from *Proceedings of the Aristotelian Society* 74: 107–29)
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: 213–25
- Statman, D. (ed.) 1993. *Moral Luck*. Albany: State University of New York Press
- Stevenson, C. L. 1944. *Ethics and Language*. New Haven: Yale University Press
- Stewart, J. A. 1892. *Notes on the Nicomachean Ethics of Aristotle*, Vol. I. Oxford: Clarendon Press
- Stinton, T. C. W. 1975. 'Hamartia in Aristotle and Greek tragedy', *Classical Quarterly* 25: 221–54
- Stock, G. 1915. *Magna Moralia*, in D. Ross (ed.), *The Works of Aristotle Translated into English*, Vol. XIX (Oxford: Oxford University Press, 1915)
- 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 (ed.) 2003, pp. 72–93. (Reprinted from *Proceedings of the British Academy* 48: 187–211)
- Susemihl, F. 1883. *Eudemii 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: 335–44
- Urmson, J. O. 1987. *Aristotle's Ethics*. Oxford: Basil Blackwell
- Walker, M. U. 1993. 'Moral luck and the virtues of impure agency', in Statman (ed.) 1993, pp. 235–50

- Wallace, J. D. 1974. 'Excellence and merit', *Philosophical Review* 83: 182–99
- Wallace, R. J. 1994. *Responsibility and the Moral Sentiments*. Cambridge, MA, London: Harvard University Press
- Watson, G. 1975. 'Free agency', in Watson (ed.) 2003, pp. 337–51. (Reprinted from *Journal of Philosophy* 72: 205–20)
1987. 'Responsibility and the limits of evil: variations on a Strawsonian theme', in F. Schoeman (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology* (Cambridge: Cambridge University Press, 1987), pp. 256–86
1996. 'Two faces of responsibility', *Philosophical Topics* 24: 227–48
- Watson, G. (ed.) 2003. *Free Will*, 2nd edn. New York: Oxford University Press
- Whiting, J. 1989. 'Comments on Susan Suavé's "Why involuntary actions are painful"', *Southern Journal of Philosophy* 27 (Supplement): 159–67
- Wiggins, D. 1978–79. 'Weakness of will, commensurability, and the objects of deliberation and desire', in Rorty (ed.) 1980, pp. 241–66. (Reprinted from *Proceedings of the Aristotelian Society* 79: 251–77)
- Williams, B. 1990. 'Voluntary acts and responsible agents', *Oxford Journal of Legal Studies* 10: 1–10
- 1993a. 'Moral luck', in Statman (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* (Cambridge: Cambridge University Press, 1995), pp. 35–45
- Williams, C. J. F. 1985. 'Aristotle's theory of descriptions', *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 edn. Oxford: Clarendon Press
- Zanuzzi, I. 2007. 'A definição do voluntario na Ethica Eudemia II.9', *Journal of Ancient Philosophy* 1(2), online journal
- Zingano, M. 2009. 'L'acte volontaire et la théorie aristotélicienne de l'action', *Journal of Ancient Philosophy* 3(2), online journal

Index

- Ackrill, J. L., 120, 157
- action
- descriptions of, 117–18
 - and Aristotle's conception of human agency, 153
 - and the primacy claim, 157–60
 - only true descriptions relevant to voluntariness and involuntariness, 155–6
 - teleological relevant to voluntariness, 138–40
- Adkins, A. W. H., 2, 7, 41
- agent
- continent, 77
 - reason-responsive, 79–80
 - continent and incontinent
 - and violence, 94, 95–6
 - faulty
 - not reason-responsive, 80–1
 - fully mature ethical agent
 - exclusive concern with, 31, 40
 - immature, 35
 - and reason, 35–6
 - as reason-responsive, 78–9
 - incontinent, 9, 10, 77
 - and contrariety to reason, 101–2
 - as reason-responsive, 79–80
 - moral, 34
 - morally responsible according to the Strawsonian interpretation, 25–7
 - non-human, 26, 27, 33, 34, 35, 37, 82
 - and reason, 34–5
 - and violent motion, 89–92, 101
 - not reason-responsive, 82
 - non-moral, 4, 5, 24, 27
 - prohairesis, 7
 - psychopathic
 - non-moral, 4, 6, 23, 27, 34, 36, 37, 77
 - non-rational, 38
 - not reason-responsive, 81–2
 - object of behavioural manipulation, 38
 - punishable, 8
 - reason-responsive as the recipient of praise and blame, 76–82
- Alexander of Aphrodisias, 127
- animals. *See* agent: non-human
- Anscombe, G. E. M., 1
- answerability. *See* Moral responsibility: as accountability
- aretê*. *See* excellence
- attitude
- educational. *See* praise and blame: as involved in moral education
 - instrumentalist. *See* praise and blame: instrumentalist
 - moral, 32, 33
 - objective, 4, 28, 29, 36
 - prospective, 29
 - reactive, 40
 - negative reactive attitudes as more basic, 48–9
 - negative reactive attitudes as sanctioning dispositions, 49–50
 - retrospective. *See* praise and blame: retrospective
- Austin, J. L., 11, 123
- backward-looking praise and blame. *See* praise and blame: retrospective
- bia*. *See* violence
- blame
- Aristotelian blame as distinct from social reproach, 42
- chastisement, 50
- childhood. *See* agent: immature
- coercion
- and hypothetical necessity, 121–3
 - non-substantive, 114–16
 - and mere instrumental rationality, 116–17
 - assumed in *EN* III 1, 134–5
 - distinct from mere instrumental rationality, 137–8

- coercion (*cont.*)
 objective distinct from subjective, 117
 overwhelming
 and the example of Alcmeon, 145–7
 as a case of violence, 132–3
 as distinct from psychological, 129–32
 as excuse, 147
 rational
 and the rationality of the agent in the *EN*, 135–7
 as justification, 140–3
 rational distinct from overwhelming, 126, 134
 compulsion. *See* violence
 continence. *See* agent: continent
- desert, 3, 33, 43
 and accountability, 50, 59
 not part of Aristotelian praise and blame, 50–1
- determinism, 129
 Diogenes Laertius, 169
 Dodds, E. R., 64
 Dover, K. J., 63
- ethical disposition, 26, 41, 43
 and pain and pleasure, 181–3
 as object of praise and blame, 68–70
 as revealed in *prohairesis*, 60
 other-regarding, 41
 perfect, 8, 9
 manifestation of, 8
 self-regarding, 41
 voluntary acquisition of, 9–10
- êthos*. *See* ethical disposition
eudaimonia, 1, 21, 52
 excellence, 1, 41
 causes us to adopt the right end, 59–60
 traditional conception of, 2
 exempting condition, 14, 27, 28
- factual error
 and different kinds of mistaken belief, 149–50
 as distinct from unawareness of the universal, 148
 as epistemic, not psychological, 150–4
 culpable, 160–72
 conservative interpretation, 162–3
 list of circumstances, 148
 relative to action-descriptions. *See* action: descriptions of
 ‘through’ and ‘in’ error in the *EE*, 163–6
 ‘through’ and ‘in’ error in the *EN*, 167–8
- Feinberg, J., 56, 62, 63, 64, 132, 136, 193
 Fischer, J. M. and Ravizza, M., 170
 force. *See* violence
- forward-looking praise and blame. *See* praise and blame: prospective
 Furley, D. J., 9, 108
- Gorgias of Leontini, 107, 109, 110
- happiness. *See* *eudaimonia*
 Hardie, W. F. R., 173
 Hart, H. L. A. and Honoré, T., 170
 Heinaman, R., 119, 162
hexeis. *See* ethical disposition
- ignorance. *See* factual error
 incontinence. *See* agent: incontinent
 instrumentality argument, 4, 33
 does not apply to moral education, 37–40
 Irwin, T., 5, 24, 27
- Joachim, H. H., 136
- Kenny, A., 113, 148, 161
- Mele, A., 130
 merit. *See* Desert
 moral demand, 33, 40, 42, 43
 moral education, 10
 as distinct from conditioning, 36–8, 39–40
 moral evaluation. *See* praise and blame: moral
 moral justification, 14. *See also* coercion:
 rational: as justification
 as different from excuse, 123–4
 moral luck, 171–2
 moral responsibility, 2, 20
 and moral assessment, 3
 aretaic perspective of, 7
 Aristotle’s own conception of, 7
 as accountability, 3, 7, 9, 20, 55
 and the Strawsonian conception, 48–50
 as interpretation of Aristotle’s concerns with voluntariness, 55–6
 as attributability, 7, 56–9
 Aristotle’s own conception of, 63, 82
 as the right interpretation of Aristotle’s concern with voluntariness, 59–62
 as ethical ascription, 62
 Kantian conception of, 2
 prospective-conditioning interpretation of
 Aristotle, 5, 30–1, 37, 77
 Strawsonian conception of, 22–3, 25
 as interpretation of Aristotle, 5–6, 22–32, 76–7
- Nagel, T., 171
 Nielsen, K., 143, 144
 Nussbaum, M. C., 11

- Owen, G. E. L., 70
- pain, 10
 and coercion, 125
 and pleasure
 as moral attitudes, 14, 183–6
 and the pain-problem for violence, 96–8
 not commensurable with wrongness, 126, 142, 146
 understood as moral pain, 98–9
 as affecting the voluntary status of the agent, 174–8
 wide notion of, 97
- praise
 not an external good, 51
- praise and blame, 20
 and moral responsibility, 3
 Aristotelian as different from the reactive attitudes, 47
 as affected by moral pleasure and pain, 189–91
 as involved in moral education 4, 70. *See also* moral education
 as *logoi*, 45–7, 62–3
 as pertinent to skills, 51–4
 as reactive attitudes, 9, 22
 as reactive attitudes according to the Strawsonian interpretation, 24–5
 conditions of, 8, 72–3
 focal and derivative use of, 7, 9, 70–2
 forward-looking. *See* prospective
 from the perspective of the virtuous spectator, 74–6, 117
- instrumentalist, 3–4, 29, 32, 33
 as distinct from prospective, 30
- moral, 33, 34
 overtly expressed for Aristotle, 63–4
 pertinent to the domain of skills, 61–2
 prospective, 4, 5, 29, 30, 31, 32, 33, 38
 recipients of. *See* agent: reason-responsive as the recipient of praise and blame
- retrospective, 32, 33, 40
 causal sense, 6, 32, 34
- virtue and vice as the objects of 43–5. *See also* ethical disposition: as object of praise and blame
- prohairesis*, 8, 9, 26, 27
 definition of, 59–60
 not present in animals, 34
 ordinary notion as voluntary adoption of ends, 120
- psychopath. *See* agent: psychopathic
- rational choice. *See* *prohairesis*
- regret. *See also* pain
 the ethical significance of, 187–8
- Roberts, J., 5, 30, 31, 37
- Ryle, G., 74
- Sauvé Meyer, S., 5–6, 24–5, 27, 28–9, 31–2, 33, 34–5, 36, 37, 40, 51, 53, 77, 109, 168, 175, 179
- shame
 and the example of the tyrant, 143
 Aristotle's conception of, 65–8
 as distinct from guilt, 64–5
- skills. *See* praise and blame: as pertinent to skills
- Smart, J. C., 3
- soul
 parts of, 77–80
- Stevenson, C. L., 3
- Strawson, P. F., 3, 4, 22, 23, 27, 28, 33, 38, 40, 42, 48, 58, 64
- Sugnomê*. *See* sympathy
- sympathy
 and overwhelming impulses, 130–2
 and the example of matricide, 143–4
 as entailing involuntariness, 145
- violence, 14
 and elemental motion, 88–9
 and pain, 92–4
 as applied to reason-responsive agents, 105–6
 definitions of, 88, 89, 94, 99, 103
 robust and core notions of, 103–5
 robust notion
 interpretation of, 178
 the definitions of, 105
- virtue. *See* excellence
- voluntariness
 and action-descriptions. *See* action: descriptions
 and involuntariness as contradictories, 96
 and involuntariness as 'willingness' and 'unwillingness?', 173–4
 and non-human animals, 11, 12
 and praise and blame, 20–2, 23, 162
 and what depends on us, 103–4, 124
 and the example of the cargo, 140
 as applied to motivational states, 128–9
 lexical ambiguity, 197–8
 Aristotle's concern with, 56, 72–3
 as openness to assessment, 12
 its significance for Aristotle's Ethics, 21, 24
 negative account of, 10
 'voluntary', 'involuntary' and 'non-voluntary', 174–8
- voluntary agency. *See* voluntariness
- Watson, G., 6, 7, 50, 56, 57, 193
- Williams, B., 64
- Woods, M., 123

