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?
>
> <div align="right">G. E. M. Anscombe, 'Modern moral philosophy', 381</div>

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* 1111b9), nor can incontinent agents as such,[10] then ethical ascription does not require that the agent act on *prohairesis* in order for his act to be ascribed to him in 'V'-terms.

The conditions of ethical ascription involve (i) and (ii). But before commenting on these conditions, it is important to rule out another condition that may be thought to be relevant to the concept of moral responsibility, and which some scholars have regarded as central to Aristotle's account of voluntariness. Some scholars, most notably David J. Furley, have argued that Aristotle's argument for the voluntary acquisition of ethical dispositions in *EN* III 5 plays a central role in Aristotle's theory of responsibility: '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* 1152a15–19), 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 *in*voluntariness. 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* 1111a22–4). The conditions of voluntariness, it seems, can be regarded as the negations of the conditions of involuntariness.