Classical Quarterly **32** (i) 84–88 (1982) *Printed in Great Britain*

A PYRRHIC VICTORY: *GORGIAS* 474b–477a[1]

Crime pays, says Polus at *Gorgias* 473. Socrates, on the other hand, maintains two propositions in the face of universal opinion:

S1. *Doing wrong is worse (for the agent) than suffering it.*

S2. *Wrongdoers who escape punishment are more wretched than those who do not.*

The argument falls into two parts: the first, up to 475e, is designed to show S1; the second, to 477a, is designed to show S2.

I

Polus will not admit S1. But he is prepared to make a concession:

A. Doing wrong is more shameful than suffering it.

Yet he maintains that the shamefulness of doing wrong does not make it bad. By 'bad' Polus clearly means 'bad for the agent'.[2] As for 'shameful', Polus is offering here the conventional view that wrongdoing is shameful, and he has no need to specify who feels the shame, or why it is described as shameful. During the course of the argument, however, Socrates will force him to specify, and hence to destroy his own position.

The argument proceeds; and Socrates wins Polus' agreement to the following:

B. We call things fine (kalon) *according as they confer pleasure or benefit or both.*

C. Thus we define the fine in terms of pleasure or good.[3]

D. We define the shameful (aischron) *in terms of pain or harm.*[4]

E. Therefore, of two fine things, that one is finer which exceeds the other in pleasure or benefit or both. Similarly of two shameful things, that one is more shameful which exceeds the other in pain or harm or both.

The principle established here looms large for the remainder of the dialogue.

F. Therefore, if A, doing wrong must exceed suffering wrong in pain or in harm or in both (475b).

G. But there is no pain involved in doing wrong.

H. Therefore doing wrong cannot exceed in pain or in both criteria.

I. Therefore, it must be more shameful in respect of its harmfulness.

J. Therefore doing wrong is worse (for the agent) than suffering wrong (*S*1).

This conclusion depends upon a dubious principle (E):

Of two fine things, that one is finer which exceeds the other in pleasure, or benefit, or both. Similarly of two shameful things, that one is more shameful which exceeds the other in pain, or in harm, or in both.

[1] This article brings together a series of arguments in my *Plato on Punishment* (University of California Press, 1981). I should like to thank Myles Burnyeat, Melanie Johnson, William Jordan and Malcolm Schofield.

[2] We should beware of Plato's tendency to present *agathon* or *kakon* without qualification or explanation of who is benefited or harmed.

[3] There is a shift here from 'benefit' to 'good'. This is reasonable, so long as Plato sticks to the prudential usage of 'good'. Cf. E. R. Dodds, *Plato: Gorgias*, pp. 249–51; G. Vlastos, 'Was Polus Refuted?' *A. J. Ph.* (1967), 455 n. 3; T. Irwin, *Plato: Gorgias*, pp. 154 ff.

[4] To avoid an illicit conversion, this principle must be derived from an induction parallel to that establishing C. Thus *aischron/kalon*, painful/pleasant and harm/benefit must be pairs of contraries.

It might be argued that here Plato commits some kind of naturalistic fallacy, so that E is false. To a deontologist, indeed, Socrates' attempt to analyse moral terms may be suspect. But Socrates does not specify who it is who derives the benefit or suffers the harm, or who experiences the pleasure or the pain. Thus the principle E lacks interpretation, and it could be filled out in such a way as to please various kinds of teleologist (a utilitarian, for example) or even someone who wishes to appeal to some *prima facie* duty to refrain from harming others. Consequently, the lack of specification is a strength, not a weakness, of the principle, rendering it adaptable to various moral positions.

In this argument, as will appear, the specification is given in terms of self-interest. The grim egoism of this view may be objectionable to us. But it is entirely consonant with Greek values, where teleological and egoistic attitudes were well entrenched.[5] So an egoistic interpretation of E would be familiar to a Greek, and so to Polus.

The specification of E, however, continues to cause difficulties. The commentators have objected that Plato equivocates, and arrives at E by improper means; and that consequently the argument is invalidated. Thus Dodds claims that Polus' concession that doing wrong is worse than suffering it 'clearly meant that it was less *ōphelimon for the community*, and from this it does not follow that it is less *ōphelimon* for the agent...'.[6] Consequently, it has been thought, the principle becomes suspect. So the apparently common-sense move at G 'there is no pain involved in doing wrong' is invalid – since the community may well find its wrongdoers painful.

Vlastos has argued that, in the rapid induction of E, 474d–475a, Socrates has failed to set up a powerful version of E.[7] All the examples (shapes, colours, sounds, music, laws and practices, the things we learn) are cited on the model of the first:

> Beautiful bodies are called beautiful either according to their usefulness...or according to some pleasure, if they make the observers feel some pleasure in observing them (474d 5–9)

Hereafter no mention is made of *who* feels the pleasure or receives the benefit; accordingly, argues Vlastos, the specification that it should be the *observer* must be maintained. Thus E becomes innocuous to Polus, who may argue at G that wrongdoing does indeed cause pain to the observer.

Alternatively, it might be claimed that the induction of E relies upon a series of examples which are not symmetrical. In some cases the pleasure of what is fine is experienced by the observer (colours, shapes, etc.), in others by the perpetrator (learning, etc.), and in others still by the subject (institutions). Plato is not entitled on this basis, therefore, to a general principle which supposes that the shameful is defined in terms of what is painful or harmful to *the perpetrator and to him alone*. Consequently, that there is no pain involved in doing wrong cannot be inferred from the earlier moves.

Now certainly both commentators are right that Plato does not specify in the principle who enjoys the benefit or the pleasure; and he does derive the principle from asymmetrical examples. But let us consider what the principle is meant to tell us. It is intended to fill out some sense of *kalon* which will enable us to analyse its meaning in its various occurrences.[8] To this end, Plato takes a series of examples – to give him the benefit of the doubt, random examples – of things that are vulgarly called *kalon*.

[5] cf. my 'The tears of Chryses: Retaliation in the *Iliad*', *Philosophy and Literature* (1978), pp. 3–22; *Plato on Punishment*, chs. 6–8.

[6] op. cit. p. 249; cf. Gerasimos Santas, *Socrates*, p. 239.

[7] op. cit. He is supported by Santas, op. cit. pp. 233 ff.

[8] Hence the insistence, throughout the passage, on our *calling* something *kalon* – 474d 5, d 6, d 9, e 3, moving to *defining to kalon* at 475a 3.

From these examples he argues that there are two common factors that appear when we speak of a thing's being *kalon* – either it is useful or it is pleasant, or both. The examples make it clear, however, that the principle thus derived is not so specific as to designate *who* (i.e. observers, perpetrators, etc.) finds these *kala* useful or pleasant. Of a *kalon* picture, it will probably be the observer who exclaims 'I find that *kalon*'. On being asked to explain why he finds it *kalon*, he will reply that it gives him pleasure. Of a *kalon* political institution, however, we might expect the exclamation to come from someone subject to it; and he will explain that he finds it *kalon* because it benefits him. The restriction, therefore, upon the principle thus derived will be that it must be the person who finds a thing *kalon* who will enjoy the pleasure or receive the benefit.[9] We may cavil at this notion that evaluative terms can be atomized in this way – this, I suspect, is the fundamental intuitive objection to the present passage. But we cannot, yet, criticize Plato on the grounds of his logic – quite properly, he refuses to specify, in the generalized principle, that the pleasure or benefit should be enjoyed only by the observer, or only by the perpetrator, or only by the subject.

But then how is the principle exploited? Polus agrees that doing wrong is more shameful than suffering it; and this, we may suppose, is the universal opinion.[10] This view of wrongdoing, therefore, will be shared by those who observe it, those who suffer it, and those who do it. In each case, they will say 'this action is *aischron*'. *Aischron*, according to the principle, is to be understood in terms or harm or pain, or both. So who suffers? The observer of wrongdoing might well agree with Vlastos, that he finds wrongdoing shameful because it causes him pain (the pain of seeing others victimized). The victim will agree, for he directly suffers the pain; and he might add that suffering wrong is bad for him, into the bargain. But the wrongdoer himself, when he, conventional man that he is, admits that his action is shameful, will disagree. 'How can you suppose that for me it is painful?' he will demand – 'why should I do it if it were?'. Plato then asks him why he still concedes that his action is shameful; and the wrongdoer, to his consternation, will realize that it is shameful for him because it is harmful for him. Now this conclusion rests upon two assumptions. Firstly, it relies upon the principle that something, called *aischron* by someone, will be so called as a reflection of the pain or the harm they are liable to suffer from it. This principle may be simplistic, but it does follow, without outrageous logic, from the examples that Plato offers. Secondly, it rests upon the assumption (conceded by Polus) that the wrongdoer will subscribe to the universal opinion that his wrongdoing is shameful; and upon the understanding that such a man neither suffers now, nor anticipates pain from his action – such as the pain of subsequent remorse.

Thus far, therefore, the objections of the two commentators are seen to be groundless. Plato's failure to specify who suffers the pain or the harm is quite proper

[9] Irwin (op. cit. pp. 157 ff.) would object that because it is egoistic, this is implausible as an account of what we value. He would prefer, for example, 'anyone – onlooker or not, truly judges that x is beautiful if x gives pleasure *to someone*' (my italics). But, as Irwin himself observes, the argument that would then follow is invalid (on much the same lines as Vlastos urges). So is the valid version of the argument based on an implausible view of our evaluative terms? I suggest not. The egoistic version of E will explain *why* we come to call things *kalon* or *aischron*; and it will be particularly plausible to a Greek, even more so to Plato. Thus Plato is explaining why we recommend things, and avoiding the dangerous waters of sympathy for others. At the same time, we may explain how an object of an action comes to be generally accepted as *kalon*, by consideration of the opinions of all those who would call it so, from all their many perspectives. That this universalizing account is the one to be given here is suggested by Plato's treatment of the exception to the universal rule, the criminal who resists any favourable characterization of his punishment. See below.

[10] cf. *tous allous anthropous*, 474b3.

in view of the fact that many people, occupying different roles, may call something *aischron* – they may observe, do or suffer.

However, there is a proviso. This reference to who is concerned, once established, should not be tampered with. It must be the person who speaks of *aischra* and *kala* who enjoys or suffers. So while the specification may vary from instance to instance, it should not be shifted within a particular example. In other words, it does not follow, from the fact that I find something *aischron*, that it is harmful *for you*.

But the first argument, to show that wrongdoing is worse than suffering wrong, can be understood without accusing Plato of shifting the specification. At 475e, Polus does indeed stand refuted.

II

Trouble looms, however, in the second part of the argument. Here Plato attempts to combine his analysis of *kalon*, principle E, with a second principle, 'the interconnection of the modalities of correlates'.[11] This combination is lethal to his argument to show S2.

The modal principle is argued for from 476b4 ff.

1. *To every agent there corresponds a patient.*
2. *The patient will be such as the agent makes it.*

Thus, as for any transitive verb, if someone hits, something is hit.

Modal principle. The effect on the patient may be qualified in exactly the same way as the act may be qualified.

If someone hits violently, something is hit violently. For straightforward examples, this principle holds. It is highly suspect, however, when the adverb is evaluative, relational or psychological. For example, 'I did this to you well' (describing my performance); does it follow that 'you suffered it well' (describing your forbearance)? If I hit you voluntarily, were you hit voluntarily?[12] We should beware of too ready a shift from active to passive.

The argument then turns to the issue of punishment.

1. *The man who punishes rightly* (orthōs) *punishes justly.*
2. *So the punisher does just things, and the man punished suffers just things* (from the modal principle).
3. *All just things*, qua *just, are fine* (476 b 1).

This move paves the way to the renewed use of E:

E. Of two fine things, that one is finer which exceeds the other in pleasure, or benefit, or both. Similarly, of two shameful things, that one is more shameful which exceeds the other in pain or in harm or in both.

4. *It would be absurd to suppose that he who suffers a just, and so fine, punishment, enjoys a pleasure.*[13]
5. *So the just punishment is fine because it confers a benefit.*
6. *Therefore the victim of punishment benefits from it.*

This argument is superficially sound up to the move from 4 to 6. For until that point, the shift from active to passive is not illegitimately exploited. The modal principle justifies 2, the shift from punishing justly to being punished justly. In this case the punisher will declare 'I am doing a *kalon* thing'. He may explicate this in terms of the pleasure just actions give him; and we will acknowledge his account of why he

[11] Dodds's term, op. cit. p. 251.

[12] cf. Aristotle, *Eth. Nic.* 1035a 15 ff.

[13] That punishments are intended to be painful to the punished is, of course, true by definition.

finds his actions *kalon*. But then the shift to the passive, to be legitimate, must carry this specification with it. Thus the man punished will be punished justly and so *kalōs inasmuch as* his punishment will give pleasure *to his judge*. It does not follow from this that he himself will benefit. We could only draw that conclusion if he too said 'this action (my being punished) is *kalon*, and I judge it so because it will give me (pleasure or) benefit'. Now Plato could derive this conclusion in the same way as he terminated the first argument, by relying upon vulgar opinion (which includes the opinion of the wrongdoers themselves) to subscribe to the evaluation in question. Thus were he able to say that the man punished, like the rest of us, calls his punishment *kalon*, then, on the grounds of logic, even if counter-intuitively, we could argue that punishment benefits the criminal. However, Plato does not do this, but instead feels the need to call upon the modal principle. He does so because he rightly doubts that the man punished too would describe what he suffers as *kalon*. In order that the punishment should be described as *kalon* at all, and for the argument to get under way, Plato must appeal to the opinion of the judge, and must judge the issue by means of the shift to the passive.

Here, then, is at last the illegitimacy of the argument. The combination of the analysis of *kalon* with the modal principle will only produce a valid argument if the reference of *kalon* is preserved in the shift from active to passive. *Kalon* relates, on the account I have offered, to whoever says the thing is *kalon* and therefore considers himself to enjoy the pleasure or the benefit. But Plato fails to preserve the specification; and so he does not show that the wrongdoer who escapes punishment is more wretched than he who does not.

New Hall, Cambridge MARY MARGARET MACKENZIE