

PLATONIC JUSTICE AND WHAT WE MEAN BY 'JUSTICE'

In the second half of Book I of the *Republic*, Thrasymachus throws down a challenge to Socrates, arguing that those who are unjust, if they are powerful enough and smart enough to 'get away with' their injustice, are always happier than the just. This challenge is reinforced by Glaucon and Adeimantus (Plato's own older brothers) at the beginning of Book II, not because they agree that the (successfully) unjust are happier, but because they want to see Socrates respond more convincingly to Thrasymachus' challenge. In the rest of the *Republic*, Socrates undertakes a response to that challenge – though the response contains many apparent digressions – arguing that the just are after all happier than the (successfully) unjust, even if we take away from the just all the rewards of reputation.¹

I. SACHS'S CHARGE OF FALLACY IN SOCRATES' RESPONSE TO THRASYMACHUS' CHALLENGE

In David Sachs's classic 1963 article, 'A Fallacy in Plato's *Republic*,' Sachs finds Plato guilty of what he calls a 'fallacy of irrelevance' in the response he has Socrates give to the Thrasymachean challenge. What this fallacy consists of, in brief, is (as I shall put it) Socrates' 'changing the subject' on Thrasymachus, Glaucon, and Adeimantus. *Their* challenge is a challenge to justice on the ordinary conception of justice, which Sachs sees as consisting in 'the non-performance of acts of certain kinds.'² (Though he does not say so explicitly, it is clear from other remarks in Sachs that he thinks of this non-performance conception of justice as identical with the ordinary *moral* conception of justice.³ Sachs calls this the 'vulgar' conception of justice, as though it is the conception which most ordinary people work with).⁴ But what Socrates does in response is to introduce a *new* conception of justice – one which would not have occurred to Thrasymachus, Glaucon, or Adeimantus in a million years unless Socrates had proposed it, and which probably would not even have occurred to us if we had not read the *Republic*. Socrates then argues – employing that *new* conception of justice – that the just are always happier than the unjust.

The new conception Socrates uses is a 'parts of the soul' conception of justice, according to which justice consists in a certain well-adjustment of the rational and irrational parts of our soul to each other. (And it certainly must be granted that this well-adjustment conception represents a quite different conception of what justice is from the ordinary moral conception, if that is put in terms of the non-performance of acts of certain kinds).

Thus, according to Sachs, the way Plato argues in Books II-IV of the *Republic* is as follows:

- 1 The psychologically well-adjusted are always happier than the psychologically ill-adjusted, no matter what actions the latter 'get away with;'

therefore

- 2 the just are always happier than the unjust;

so we may conclude that

- 3 Thrasymachus, Glaucon, and Adeimantus are quite wrong to suppose that the unjust, when sufficiently successful at ‘getting away with’ their injustice, are happier than the just.

Sachs’s criticism of this argument is, in effect, the following:

If this is to be a good argument, then the conception of justice involved when one infers (2) from (1) must be the same as the conception of justice involved when one infers (3) from (2); but the former conception is the Platonic conception, while the latter conception is the vulgar conception. So the argument from (1) to (2) is irrelevant to the desired conclusion (3).

Another way of representing what Sachs here calls a fallacy of irrelevance is as a fallacy of equivocation. The idea is that Thrasymachus means one thing by ‘justice,’ Socrates another. So if Socrates, in saying he *disagrees* with Thrasymachus and that he has an argument against him, does not employ the meaning Thrasymachus employs, he will be equivocating in his step from (1) and (2) to (3). In spite of what he says, Socrates will not even be disagreeing with Thrasymachus. What Socrates *denies* (about psychological well-adjustment) is not what Thrasymachus *affirms* (about the ‘non-performance of acts of certain kinds’).

As Sachs puts the issue, it is one of a failure of match between the conception with which Plato proves the greater happiness of the just and the conception which Thrasymachus, Glaucon, and Adeimantus employ in their challenge. At the end of the first section of his paper, Sachs puts it in this way:

In what follows, I argue that Plato failed to relate the two conceptions of justice adequately, and that it is implausible to suppose that the omission, a complex one, can be remedied. Consequently, Plato’s conclusions about happiness and justice – as he conceives the latter – prove irrelevant to the dispute between Socrates and Thrasymachus (and Glaucon and Adeimantus, in so far as they, too, are concerned with the happiness of vulgarly just and unjust men).⁵

The suggestion here is that Plato has turned his back on the question what it is that is really being talked about in Thrasymachus’ challenge. (From now on, I shall understand ‘Thrasymachus’ challenge’ as short for ‘the challenge of Thrasymachus, Glaucon, and Adeimantus’).

II. STRATEGY OF MY RESPONSE TO SACHS’S CHARGE OF FALLACY

It will be my aim in the present paper to show that Sachs's charge is entirely misguided, and should be dropped. I shall argue that Sachs distorts what is going on in the *Republic* because of a false – and certainly un-Platonic – doctrine, to the effect that meaning (or conception) determines reference.⁶ That is, Sachs supposes that when Thrasymachus, Glaucon, Adeimantus, or Socrates say something about justice, the justice each of them is talking about is determined by the particular conception of justice each is working with in what he says. Since the conceptions involved (what these individuals *mean* by 'justice' or *dikaionê*) are so different – that of Thrasymachus, Glaucon, and Adeimantus being (supposedly) a moral conception (thought of in terms of the non-performance of acts of certain kinds), that of Socrates being a purely psychological conception – it will follow pretty directly that, though they are using the same *word*, they are talking about quite different things.⁷ From this originates the charge that in replying to the challenge of Thrasymachus (as amended by Glaucon, and Adeimantus), Socrates is changing the subject.

In my view, this supposition that meaning (conception, concept, linguistic intuitions)⁸ determines what we are talking about, has the following disastrous effect: it invites us to confuse what 'our' conception of justice *determines* (in the world) with *what justice really is* – with the real nature of justice. What that 'real nature of justice' is, I shall maintain, it is independently of how we may conceive of it.

To see this, consider a parallel case. Consider the real nature of what Greeks before Hippocrates called 'the Sacred Disease.' As this name implies, at least some pre-Hippocratic Greeks thought of this 'Sacred Disease' in terms of possession by something divine or something *daimonion* – in modern dress, possession by evil spirits. 'The Sacred Disease' (I shall suppose) picks out the brain disorder epilepsy; and it picks out that brain disorder *however* Greeks prior to Hippocrates may have conceived of the relevant seizures -- e.g., in terms of possession by evil spirits or devils. What is more, it is, in general, the *intention* of such pre-Hippocratic Greeks that they be *understood* as referring to that brain disorder (if that's what it is), even if they happen to have the mistaken belief that this particular disorder is to be identified with possession by evil spirits. When a child is having a seizure and the parents send for help, what they want help with is *whatever the disorder really is*: possession by evil spirits, if that's what it is, a brain disorder if *that's* what it is. ('Whatever the disorder really is' is a kind of blank check issued by the parents. But of course it is a blank check that is only to be filled out with what the actual thing is. For someone else to fill it out *ad lib* is to violate the parents' intentions in issuing the blank check).

In the view I am urging, then, what Justice *is* should not be identified with *our conception of Justice* any more than what the Sacred Disease is should be identified with our conception of the Sacred Disease. More precisely, the real natures of Justice and the Sacred Disease are not to be identified with what the speaker's conceptions of Justice or of the sacred Disease *determine* (in the world). (As if what they are talking about could only be something that corresponded fully to those conceptions). On the contrary, it is the

real natures of Justice or the Sacred Disease that determine, at any rate in favorable cases, what we are talking or thinking about.

My thesis, then, in a little more detail, will be this: If

- A we avoid the confusion of our *conceptions* of justice or the Sacred Disease with what justice or the Sacred Disease *are*;

and, more precisely, if

- B we avoid supposing that the objects we are talking about and *intend* to be talking about can only be something that fully corresponds to our conceptions (i.e., can only be what our conceptions determine);

then we'll see that

- C Plato is *not* open to the particular charge that Sachs makes to the effect that Plato 'changes the subject' on us from the moral virtue justice to a purely psychological state of soul.

For should Plato be *right* about what the real nature of justice is – *that* is the real issue here, not fallacies of irrelevance or fallacies of equivocation – then it can be argued that

- D it is that real nature of justice that Thrasymachus too meant to be talking about. Thrasymachus' intention was to speak, not of (i) justice according to his conception *whether or not justice-according-to-his-conception corresponds to the real nature of justice*, but rather of (ii) the real nature of justice *even if that is different from what he conceives it to be*.⁹

Hence, on the same assumption that Plato was right about the real nature of justice,

- E the reason Plato has not changed the subject from a *moral* virtue to a purely psychological state of soul (or virtue) is that the subject never *was* this supposed *moral* virtue. Whatever other Greeks might have thought of this question,¹⁰ it was Plato's view that the subject-matter of the *Republic* always *was* the psychological state mentioned in (D). So (3) does not refer to some justice other than the justice referred to in (1) and (2).

In the course of explaining the Platonic position in the *Republic*, I shall also discuss briefly the somewhat differing position on the real nature of justice that I find in a number of passages from Plato's early dialogues. That different position I call the *Socratic* position on justice (justice as a skill or expertise at obtaining the maximum good possible for one in the circumstances), identifying it with the position of the historical Socrates. (In the sequel, the term 'Socrates' will stand for the character in Plato's *Republic*, and hence for the Platonic position, while I shall use 'Socratic' and 'the

historical Socrates' for the historical Socrates whose position I claim to find in much of the early dialogues).¹¹ I shall argue briefly that the Platonic position (justice as a certain psychological well-adjustment) derives from that Socratic position (justice as expertise), as a result of a change I shall sketch (sec. VIII below) between Socratic and Platonic psychologies of action. I shall further claim that there is no substantial difference between the Socratic and Platonic positions on human goodness that does not flow from this difference in psychology of action. Thus, in my picture, the Plato of the *Republic* thinks, following the historical Socrates, that the virtue of Justice *is* a self-interested psychological state of the psyche that is not at all moral. What we call ethics is, for the historical Socrates, part of the science of psychology: The just or good person will, as a purely factual matter, be the person good at maximizing his or her own happiness.¹² Since I view Plato, in his theory of virtue, as departing from the historical Socrates solely in what flows from his psychology of action, I find pretty well the same conception of the good person in the *Republic*, *mutatis mutandis*. The result is that the Platonic psychological conception of justice really is as different from the moral conception held by Thrasymachus, Glaucon, and Adeimantus (not to mention most modern readers) as Sachs says it is. But that fact is irrelevant to the issue whether Plato is allowing the character Socrates to change the subject on Thrasymachus, Glaucon and Adeimantus. The only issue is: who is right about the real nature of justice, and does the real nature of justice make us happier than successful injustice would?

III. A BROADER, MORE INTUITIVE CHARGE OF 'CHANGING THE SUBJECT' – ONE THAT DOES NOT APPEAL TO MEANINGS, CONCEPTIONS, AND THE LIKE

Even if I convince my readers on the importance of rejecting the methodology of meanings, fallacies, and so forth, that Sachs, with many other modern philosophers, relies on, I will still need to broaden the look I am taking at the accusation of changing the subject. For there may be those who, while agreeing that Sachs' particular accusation is based on questionable philosophical assumptions about meanings, still feel that Socrates *has* changed the subject on Thrasymachus, Glaucon, and Adeimantus. For it remains the case that, even without Sachs-type argumentation involving meanings, conceptions, and the like, alert readers have the hardest time escaping the doubt that Plato has changed the subject on them. The feeling is as simple and uncomplicated as this:

Look, it's totally unreasonable to suppose that justice could possibly be anything like the self-interested state of psychological well-adjustment Plato says it is. What Plato says is just too distant from anything we could reasonably suppose justice to be. Surely all we can reasonably suppose justice to be is the familiar other-regarding virtue.

This is not an unreasonable expostulation. (It is doubting that the real nature of justice really *could* be a certain purely psychological state). It is saying, in effect,

Even if the real nature of justice *should* turn out to be the self-interested state of psychological well-adjustment that Plato says it is, he really does need to make at

least a *prima facie* case for that being so. And we do not even have the beginnings of such a *prima facie* case in the *Republic*.

So even after I have responded to Sachs, it will remain to offer some sort of response to the broader version of the charge, by offering a *prima facie* case for justice being as surprisingly different as Plato makes it out to be. Without this, we are in no position to affirm that what both Thrasymachus and Socrates are talking about, and *intend* to be talking about, is the real nature of justice (even if that real nature differs from what we conceive it to be). The character Socrates will still be under the suspicion of changing the subject from what justice is to some totally different topic.¹³ I turn to making that *prima facie* case, and so dealing with this broader charge, at the end of the paper (sec. VIII).

First, however, I lay out in a little more detail how the Socrates of Books II-IV of the *Republic* introduces his new parts-of-the-soul theory of what justice is (sec. IV). Once I have done this, and emphasized how, on Sachs's assumptions, the change of subject *would* indeed be flagrant once one notices it, I turn (sec. V) to showing that Sachs thinks it irrelevant to the charge he brings *what* the real nature of justice might be. What Socrates must respond to, Sachs thinks, is the challenge put in terms of *Thrasymachus'* conception of justice – which tells us that Sachs would be unimpressed by the notion that before we can decide whether Thrasymachus is right in his claim that (successful) injustice makes us happier than justice, we first need to know what justice is. In sec. VI, I show how behind Sachs's charge lies a certain theory of the identity of questions (and so of the nature of disagreement) that is popular among philosophers, especially modern analytical philosophers, and which has been given classic expression in the opening paragraphs of Frege's great paper, 'Negation' of 1918. In sec. VII, I explain why I reject this theory of the identity of questions (and also believe Plato would have rejected it). In sec. VIII, I respond to the broader charge of changing the subject just introduced, in so doing laying out the Socratic theory of what justice is from which I claim the Platonic theory derives. Then, in a concluding section, I offer first a summary, then some concluding reflections on how I think philosophers came to embrace the convenient, but I believe mistaken, notion of meaning that so badly misleads Sachs and other modern interpreters of the *Republic*.

IV. THE NEW THEORY OF JUSTICE IN *REPUBLIC* II-IV; HOW THE CHARGE OF 'CHANGING THE SUBJECT' MAY SEEM INTENSIFIED WHEN WE NOTICE THAT MOST CITIZENS OF THE JUST CITY ARE NOT THEMSELVES JUST.

At the beginning of Book II of the *Republic*, Glaucon lays out a social contract theory of justice, according to which one can be sure one does well in acting justly solely *when one has an agreement in place that all will act justly*. Just deeds by themselves – without an agreement – are often disadvantageous to me. True, those disadvantages may be offset by advantages that accrue if others act justly towards me. But such advantages can be reliably obtained only if all the people involved together accept, and are induced to abide by, a certain sort of agreement to act justly. Once we have such a justice-by-agreement in place, the only net losers in justice-by-agreement are those who, with or

without agreement, are powerful enough to get away with their injustice, or who are clever enough to succeed in concealing their injustice from the notice of others. (The idea is that the seeking of happiness, like the seeking of property, is fundamentally a zero-sum game, so that if others ‘get more’ than you, you get less. Most people not only lack the power to be happy without a social contract, they are much less happy because others are all free to exploit them. For them, a social contract is the best option).

What is important from our point of view about this version of social contract theory is that it implies that, but for the rewards of reputation and the advantages of equalization that the social contract brings to all but the most powerful, justice would not be rewarding. This view Glaucon expresses, prior to his social contract theory of justice-by-enforced-agreement, as the view that we are not made happy by justice *in itself*, but only by the consequences of good reputation from justice. Roughly, what justice would bring to us ‘in itself’ is what it brings us prior to the social contract and setting aside any rewards of a reputation for being just.¹⁴ One contracts with others to avoid the bad effects of the injustice of others.

Socrates leaps to the defense of justice as advantageous *in itself*. He begins by asking what justice is. First, he claims that

R1 what justice is (in a person) can be best observed by seeing what justice is in a certain sort of ideal city (368c-369a).¹⁵

If we construct such an ideal city in our imagination, we see that

R2 justice in that city will consist in a certain ordering of that city by virtue of which each of its three classes – (a) the intellectual class (which rules, with knowledge, in the interest of the whole: 428c11-d3, cf also 519e1-520a4, 420b5-7, c1-4, 421b3-c6, 465e4-466a6), (b) the military class (which defends the state externally, and keeps order internally in obedience to the true beliefs about law implanted in them by the philosopher-rulers: 429c5-d2, 430a1-b5, 433c7-8), and (c) the artisan class (sticking to their agriculture, carpentry, and so forth, and agreeing as to who should rule: 431d9-e6, 432a7-9, 433c6-7) – ‘does its own’ (i.e. does its own work, carries out its own function)¹⁶ and does not interfere with the members of other classes doing *their* own work (433a8-b1, 434b9-c5).

Applying this structural account of justice in the city to the individual psyche, we get that

R3 this same structure, justice, when it shows up in an individual psyche, consists in an ordering of the three parts of the psyche – (a) reason (which rules, with knowledge, in the interest of the whole: 441e4-5 with 442c5-8), (b) the spirited part (which reinforces the commands of reason: 441e5-6, 442c1-3), and (c) the appetitive part (agreeing with the other two parts as to who should rule, and not rebelling: 442c10-d1) – an ordering in

which each ‘does its own’ (i.e., does its own work, carries out its own function) and does not interfere with others doing *their* own. Injustice is a corresponding disorder of the psyche (443b1-5, 444a10-b8).

With this account of justice in hand, then – making justice in the psyche structurally identical with the justice of the ideal city, so that justice may now be said precisely to *be* the structure identical in both city and individual psyche¹⁷ – Socrates is ready to argue, in two further steps, that the just person is happier than the unjust person. First, he says,

R4 We do well to compare this well-ordering in the psyche with yet another kind of internal well-ordering. This time the well-ordering is within the body: that well-ordering of naturally opposed elements of the body which we call health (444c-e).

That comparison made, we come, second, to a relatively brief argument for the greater happiness of the just person over that of the unjust person (444e7-445a4), even when the unjust person can ‘get away with’ his or her injustice. It’s this:

R5 As all the food, drink, wealth, and power in the world does not profit someone whose health has been destroyed, so too all such supposed advantages do not profit one with that disorder of the psyche that Plato tells us injustice consists in (*Republic*. IV.444e-445b, cf *Crito* 47a-48a, as well as *Gorgias* 504d-505b). Therefore Thrasymachus is wrong to think that those who act unjustly (even when they can ‘get away with it’) are happier than the just.¹⁸

Such is the (breathtaking) argument of *Republic* Books II-IV. As we see from the accounts of Justice in city and soul in (R2) and (R3), this argument is conducted entirely in terms of the supposed structural parallel between the state with each of *its* three internal parts (the three classes) doing its own, and the individual psyche with each of *its* three internal parts doing its own. From the account of justice in the city that we find in (R2), we have

JCITY justice in a city is a matter of the existence of a certain internal structure to the city, that is, the holding of certain internal relations between the parts of that city (each part does its own, and does not interfere with other parts doing *their* own);

and from the account of justice in the individual that we find in (R3), we have

JPSYCHE justice in the individual is a matter of the very same internal structure to the individual psyche – the very same internal relations between the corresponding parts of that individual’s psyche (each part does its own, and does not interfere with other parts doing *their* own).

Thus, from Sachs's point of view, this argument is conducted entirely in terms of the new Platonic conception of justice as psychological well-adjustment. So far, at any rate (though see n.27 below), Plato's argument totally ignores the vulgar (moral) conception in terms of the non-performance of acts of certain kinds.

One might try to argue in Plato's defense that since justice in the city turns out to be each part of the city doing its own and not interfering with others doing *their* own, that Plato *is* allowing for at least some reasonably close analogue to justice as non-performance of acts of certain kinds. For why wouldn't it be the case that

for each citizen to do his or her own and not interfere with others doing *their* own
is

for each to do something *like* refraining from performing acts of certain kinds?

Wouldn't this at any rate be *close* to the non-performance view of just acts?

But such a defense of Plato is both unwanted and mistaken as a reading of Plato. The *Republic* is not arguing that a just individual or even a just citizen is an individual or citizen who refrains from performing acts of certain kinds.

To see this, notice the following point about Plato's decision to study the just psyche in terms of the (ideally) just city as a model: Not every feature of the model is appealed to in drawing conclusions about justice in the psyche. For example, there is very little in the *Republic* that appeals to a city's showing justice or injustice in its external relations to other cities. This suggests that it must remain an open question whether the *Republic* supposes it can study the justice of individuals in terms of their external relations to other individuals (as in each doing his or her own and not interfering with others).

It is true that it is very tempting to suppose – and many interpreters of the *Republic* indeed *have* supposed – that Plato intends us to infer that the just citizen is the citizen of the ideal city who in his or her external relations to other citizens does his or her own and does not interfere with others doing their own. It is tempting because, first,

J1 (JCITY) analyses the justice of a city in terms of the holding of certain internal relations between the parts (presumably including the citizens) of that city in which each of the parts (and presumably each of the citizens) does its own, and does not interfere with others doing their own;

and then, second,

J2 those relations *internal to the city* are, with respect to those citizens, relations *external to the citizens* – relations in which they stand to other citizens and in which they each do their own, and do not interfere with others doing *their own*;

hence, it is natural to conclude,

J3 being citizens of a just city by virtue of each doing his or her own and not interfering with others doing their own, surely each is a just citizen.

Here we have an inference from the idea of a just city in (R2) above – an inference from

JCITY justice in a city is a matter of the existence of a certain internal structure to the city, that is, the holding of certain internal relations between the parts of that city (each part does its own, and does not interfere with other parts doing *their own*);

to the conclusion that

JCITIZEN the justice of an individual citizen is a matter of *external* relations between that citizen and other citizens (where each citizen does his or her own and does not interfere with other citizens doing their own).

There is surely something very plausible about this inference. We might ourselves – at any rate, temporarily – buy into the idea in (R2) or (JCITY) of a just city as a city with the appropriate internal relations between its parts: the three classes, and the individuals within those classes. Leaving aside the classes for the moment, we might tell ourselves that justice in a city surely *is* a matter of appropriate relations internal to the city between the individuals in that city. (Even Rawlsians will not disagree *so far*). But if we do buy into this idea, then surely we will want to see precisely those relations internal to the city as constitutive of the justice of the *citizens*. But that tells us precisely that the relations *internal* to the city are, at the level of the individuals who are citizens, relations *external* to the individuals who are citizens. And this is precisely (JCITIZEN).

But, to turn now to the text, it can be shown that this appeal to the inference from (JCITY) to (JCITIZEN), as a way of showing Plato sticking to a non-performance criterion for justice – is closed off. This point was made convincingly in sec. I of Cooper 1977. Let us go back to (R3), and

JPSYCHE justice in the individual is a matter of the very same internal structure to the individual psyche [as we find in the just city] – the very same internal relations between the corresponding parts of that individual's psyche (each part does its own, and does not interfere with other parts doing *their own*).

By this account of justice in the individual psyche, most citizens of the just city *are not individuals who are just*. For consider: A citizen of a just city (externally) does his or her own. For that citizen to be a just individual by the account in (R3) and (JPSYCHE), it would be necessary for each of his or her internal parts to do its own. That is, it would be necessary for his or her Reason to be wise and have knowledge; for the spirited part to be tenacious of the beliefs given to it by reason; and for all three parts to agree that reason is to rule. That's what justice *is* according to the *Republic*.

So will it be the case for any given citizen of the ideal city that each of his or her internal parts does its own? It will not. The only just individuals amongst the citizens of the just ideal city will be the members of the intellectual ruling class – who have a soul in which their rational part ‘does its own.’ For it is they alone whose rational parts have knowledge (Cooper cites 441e4-6 with 442c5-8). It follows that no other citizens of the just ideal city will themselves be just, since no other citizens will have the knowledge necessary for their rational part to do *its* own (cf. 443c9-d1 with n.20 below). Citizens of the two lower classes who ‘do their own,’ and do not interfere with others who are doing *their* own, will not themselves be just.¹⁹ Thus the principle (JCITIZEN) is false: the justice of the individual citizen does not consist in that citizen doing his or her own.

It is idle here to bring up a *secondary* sense of justice, called ‘demotic’ justice, for which it is sufficient that one do one’s own by obeying the orders that are the result of the knowledge not of one’s own intellect, but of the intellect of someone in the ruling class (590c2-591c3).²⁰ To think Plato is saying that *this* is justice would be for us as *interpreters* to change the subject from what Plato is talking about.

One may be tempted here to object that the textual arguments against (JCITIZEN) are overly literal, attributing to Plato what is little better than a *stipulation* concerning the just citizen. But to suppose this is what is going on is to miss the importance of raising the question what, in the model Plato is constructing (the just city), is intended to model features of the just psyche, and what is not so intended – what is a mere ‘artifact’ of the model.

Look at the situation in this way. Most moderns, I doubt not, had they realized that Plato was thus committed to denying the connection between relations internal to the state, and relations external to the citizens of the state, would have rejected the analogy between city and individual psyche. This would be especially true for those who, like Aristotle, believe in *dispositions* and *character* – as if accepting justice as an inner structuring of the parts of the psyche required rejecting

JCITIZEN* for an individual (i.e., the individual citizen of the ideal city), to be just is for that individual citizen to [be disposed to] do his or her own and not to interfere with others doing their own, regardless of what holds between the internal parts of his or her soul.

If the city-soul analogy requires giving this up, why would such interpreters not reject the analogy? Is it not practically a matter of *logic*, one can imagine them saying, that a theory of justice requires that the *internal* relations between parts of the city are identical with the *external* relations between the citizens of that city? But, in that case, if the justice of an individual is that which is had by the citizen of a just city, then, by modern lights (at any rate), there would be no need to suppose justice was a structuring of the psyche of the individual, since all we need in the soul is a [disposition to the] doing of one's own, and [to] refraining from interfering with others doing their own. At least for much modern philosophy, it would not matter (except causally) what inner states one had.

From the point of view of this alternative model, the inner psychic structure of the citizens would play no role – unless possibly causally – since with the psyche, as long as this disposition to behave appropriately in one's external relations with others remains the same, the parts need not have any particular inner psychic structure. Thus there is no inquiry, within this alternative model, into the psychic structure of the citizens, so long as they are [disposed to] do their own, and not to interfere with others fulfilling *their* own.

Now I deny that this model reflects Plato's argument in the *Republic*. As I see it, the *correct* model for the just psyche is not the alleged justice of the *citizen* of the just city, but only the justice of the *city*. It is the city, not the individual citizen of the city, that is the entity we must look to in the model Plato intends; for it is the only entity in the model the inner structure of which matches that of the psyche. The inner structure of the soul of *the citizen of the just city* is irrelevant in this alternative model, since what this model requires of him or her is simply to have the disposition to do his or her own (regardless of any inner structure in his or her psyche). My point here is this: The ideal city is less an account of the justice of an ideal city (and of its citizens) than it is a heuristic mental construction aimed at an account of justice in an individual of no matter what city. Once we have this mental construction, certain further questions may still be raised, of course. For one thing, does Plato's Callipolis also give us the kind of theory of the just city that would be of interest to political philosophy? The answer here is far from assured, since it is hard to imagine that a political philosophy would demand a theory of a just city unless its *citizens* were just. And we have seen that the citizens of Plato's Callipolis are *not* just, aside from members of the ruling class. For another, one who finds the ideal city described to be quite fascinating – and who would not? – might well wish to ask whether such a just city could ever be realized – a question on which, it is well known, Plato blew hot and cold. (Thus at 472c-e, he maintains that it doesn't matter for his purposes whether it is realizable or not, while elsewhere he longingly describes conditions under which it would be realized: see the passages cited in n.51 below. We have already noted the recapitulation of *Republic* II-V which we find at *Timaeus* 17b-19c, without the slightest indication of any interest in the structure of the psyche. But is Plato asserting this ideal city is more than an abstract possibility that he thinks might be too difficult ever to realize? Do the *Statesman* and *Laws* suggest he might be? I must leave these larger questions aside here).

The tendency for moderns to think that Plato is interested in the justice of the citizen of the just city is accentuated by the customary translation, following Cicero, of the Greek title *Peri Politeias* as ‘The Republic.’ As Schofield 2000, 228, 231, points out, a more correct translation would be ‘The Constitution.’ The translation ‘The Republic’ suggests we are examining the justice of a city and the justice of individual citizens of that city. How, on such a view of the principal subject-matter of this dialogue, could Plato not be endorsing (JCITIZEN*)? The truth is, however, that the *politeia* in question is not the city but the structure of the city. But that is, in turn, the structure which Plato is saying is *identical* with the structure of the soul. True, the word *politeia* is most often reserved for the city. But it is certainly also used for the structure of the soul (cf. 606b1-c7, 608b1). It is in any case that structure of the soul (which is what justice in the individual *is*) that I am urging is the *principal* subject-matter of the dialogue. This point is surely confirmed by 592a10-b3 at the end of Book IX, where Socrates refers to the city we have founded ‘in discussion (*logois*),’ which is laid up as a paradigm in heaven for one to see – a paradigm which, once one has seen it, one who wishes may establish *in oneself*.

What we need to see here about what Plato is doing is that Plato has a whole new idea about what justice in the individual is, based on a whole new idea about what the psyche is. (More on this below, in sec. VIII, when I get to speaking of the Socratic Psychology of Action, and of the corresponding Platonic theory of action.) According to this new idea, the parts of the psyche are analogous to certain parts (namely classes) of a certain *ideal* city. Notice that this is not simply *any* city, not even simply *any* city that one might be inclined to call a *just* city; and it is not simply *any* classes. There is no simple-minded semantical principle here going from our use of a word for a city to our use of the same word for a psyche.²¹ The city in question, and its classes, are, on the contrary, carefully constructed (to be gerrymandered even, if necessary – so that one’s account of the just city and the just soul stand in what Rawls would later call ‘reflective equilibrium’: see 434e4-435a3, 420d5-e8) in order to give the appropriate relations needed for Plato’s new theory of the psyche. So there is no automatic inference, as in (J3) above, from being a citizen of the just city to being a just individual. Instead, there is a new theory of the psyche and it is in terms of that new theory of the psyche that the ideal city gains, by a highly artificial construction, many of its characteristics.²²

Thus, what is shown by the textual arguments given above is simply that any inference from (JCITY) to (JCITIZEN) does not capture an intended feature of the model.

If, on the other hand, as Sachs insists, we were to talk about justice as *we* understand justice – including the justice of a citizen of a just city – then there *has* been a change of subject, and a really flagrant one at that.

On Plato’s propensity to change the subject – if that is what he is doing – it may also be pointed out that there is another piece of evidence lying to hand for Sachs if he wishes. For the charge may be extended backwards to the more Socratic Book I. For

Book I.343bff has its own version of challenge and reply – challenge from Thrasymachus, reply by Socrates. We should surely say here too, on the sorts of assumptions that Sachs is making, that we understand Thrasymachus’ challenge rather better than Socrates’ reply. It is natural, of course, to suppose that the challenge is to justice as we ordinarily understand it. Now, Thrasymachus arguably (n. 2 above) poses his challenge to Socrates in terms of non-performance of acts of certain kinds. But if this is the challenge, then, in his reply in Book I, Socrates seems, almost constantly, to be pushing a quite different conception of justice – justice as a *skill* or *expertise* of some sort. The situation is even more blatant with Polemarchus earlier. Polemarchus wants to characterize justice in terms of telling the truth and returning what has been entrusted to one (rendering to each what is owed), and Socrates in his response turns this justice into a skill or expertise.²³ Is this not yet again Socrates trying to get away with an irrelevant change of subject, just as Prichard and Sachs say about Books II-IV? When Thrasymachus throws down a challenge to the ordinary conception of justice – what we all *mean* by ‘justice’ or *dikaiosunê* – do we not see Socrates reply by speaking only of a certain skill or expertise? A modern reader must wonder: ‘How could we, by looking to a *skill* or an *expertise*, arrive at any idea of justice as a criterion for judging just actions? Or at any idea of the sort of moral character just persons have? The discussion seems to be about something other than what we mean by ‘justice.’ Socrates seems to be changing the subject on us.’

So Sachs’s charge, if it can be made out at all, should be extended backwards to Book I. But perhaps, one tries to reassure oneself, the occurrence of such a feeling about the irrelevance of bringing up expertise in this context is not too serious? After all, the idea of justice as nothing but a certain expertise pure and simple²⁴ does not appear again in the *Republic* after the first Book.

If one’s interest is solely in the *Republic* from Book II onwards, it may not be so serious that Plato changes the subject from Thrasymachus’ conception to a skill or expertise in Book I, and then changes the subject on *himself* as he moves from most of Book I²⁵ to Book II. Even so, on Sachs’s view, it must surely suggest in Plato a serious pattern of inability to settle on what it is that is being talked about. (This charge might then be extended to other early dialogues of Plato).

V. HOW IT IS IRRELEVANT TO SACHS’S CHARGE WHAT THE REAL NATURE OF JUSTICE IS: ALL THAT MATTERS ACCORDING TO SACHS’S CHARGE IS THRASYMACHUS’ CONCEPTION OF JUSTICE

In sec. I above, I have said that Sachs construes Thrasymachus’ challenge to Socrates as a challenge to show that justice *as conceived by Thrasymachus* will make people happier. But consider now the following question, not raised by Sachs: Suppose that Thrasymachus himself has a *wrong* conception of justice. That is, suppose Thrasymachus has *the wrong answer* to the question ‘What is justice?’ Would this, on Sachs’s argument, make the slightest difference to the charge that Socrates’ reply is irrelevant to the challenge laid down? It would not. *Even if psychological well-*

adjustment were what justice really is, it would still be the case, according to Sachs's argument, that Socrates' reply is irrelevant. This will be evident from Sachs's actual words (quoted in sec. I above) about Socrates failing to relate adequately the Platonic conception and the vulgar conception of justice. Sachs makes no reference here to the real nature of justice, and speaks only of conceptions of justice (though see n. 5 above). Thus, suppose for the sake of argument that Thrasymachus' conception of justice as 'vulgar justice' is wrong as to the real nature of justice, and indeed, that the real nature of justice is psychological well-adjustment. In those circumstances, the passage quoted in sec. I shows this would make no difference to Sachs's accusation. Sachs would not allow Plato to say, 'See, Thrasymachus was all along talking about psychological well-adjustment.' Sachs's claim is this: What Socrates has to respond to is not to a challenge that the real nature of justice is such that it does not make us happier, but to the challenge that *justice as conceived by Thrasymachus* does not make us happier.

Here we begin to see a real difference between Sachs's procedure and Plato's. *Sachs* thinks we need to determine what conception Thrasymachus is actually employing, and then – *using* that conception – assess the relevance to the challenge of any conception Plato employs in replying to that challenge. For Sachs, as we shall see, it is relatively unproblematic what these two conceptions *are*; what is difficult to see, according to Sachs, is how Socrates can connect the conception of justice *he* uses in his reply, to the conception *Thrasymachus* employs in issuing his challenge. On the other hand, *Plato* takes it as unproblematic that Thrasymachus and Socrates *are* talking about the same thing when they 'talk about justice'; what is difficult, Plato thinks, is to see exactly what that justice is.

Let me put the point about how Plato sees the problem here in a more metaphysical way – parallel to the talk of Forms that we come upon later in the *Republic*. What is difficult to see, according to Plato, is what that *real nature of justice is* that Plato takes Thrasymachus and Socrates both to be trying to talk about, and indeed to be disagreeing about. Plato thinks that all he has to have Socrates do is to show that the real nature of justice is such that it makes us happier. He doesn't think he has to worry about any conceptions Thrasymachus may have about justice (at any rate if they are mistaken).

What would Sachs say to a Platonic response to the effect that it is the real nature of justice that Thrasymachus is denying makes us happier? I suspect Sachs's reply would be something of the following sort:

Look, I don't know about this 'real nature of justice' you are talking about, Plato. All I am doing is raising a question about Socrates responding to Thrasymachus. I'm examining an *argument* Plato gives. (That's what we analytical philosophers do: we examine arguments). I am not ranging off into distant areas of metaphysics. I'm talking about a response to a challenge here. All I'm saying is that if the response to the challenge is to be *relevant* to the challenge, it must speak to the challenge actually made. Now, it's Thrasymachus' challenge. So he presumably knows what challenge he is laying down. Is that a challenge about

justice as psychological well-adjustment? Prior to Book IV of the *Republic* that conception of justice will not so much as have *occurred* to Thrasymachus.²⁶ So that can't be what the challenge is. No, it's surely obvious what the challenge is. It is a challenge to justice on the vulgar conception of it. And to that challenge, all of Plato's arguments are irrelevant.²⁷

One may feel considerable sympathy with the hard-nosed reply that I imagine Sachs to be making here. But I hope this imagined reply *does* make clear why I think that *Sachs doesn't care what the real nature of justice is, or even whether or not there is such a thing as the real nature of justice*. He's just *analyzing arguments*.

Let us now consider what Plato would say to the line of thought. What if he were to say, along the lines broached in sec. II above,

Look, I agree entirely with Sachs that Socrates must show that the justice Thrasymachus says makes us *less* happy in fact makes us happier. But I say that what Thrasymachus is talking about *is* the real nature of justice – even if what that real nature *is* happens to differ from what his conception would make it. We all know, if we are adults at all, that our conceptions of the things we talk about are bound to be inadequate in various ways to the realities we want to talk about. But it is no part of what we want to do, when we talk about something, to talk (not about what the actual things *are* that are there, but) about some fantasy-object which is what *would* exist if our conceptions were totally adequate. (Or about some object in another possible world). We want to talk about what the actual things are. So I have *not* changed the subject on Thrasymachus. In saying that justice makes us less happy, he, like everyone else, wants to be understood to be talking about the real nature of justice, even if that is different from what his conception of justice would make it. And it is that real nature of justice which I am saying is psychological well-adjustment, and which I am saying makes us happier. Of course, Sachs may object if he likes that it is I who am wrong about what the real nature of justice is. I've no doubt he thinks I *am* wrong about what it is. But *that* is not the objection he makes to me in his paper. The objection in the paper – (in effect) that I have to respond not to a claim about the real nature of justice, but to a claim about some Thrasymachean *conception* of justice – is entirely beside the point. The point *is* to determine whether *what justice really is* benefits us or not.²⁸

Plato's line of thought, here, we should note, is entirely in accord with his view that before we can know whether justice is more profitable, we must know what justice *is*. In fact, there is something decidedly odd in Sachs's idea that Plato keeps changing the subject on us, as though he can't keep his mind on what it is he is talking about. For Plato's first priority, in the *Republic* as elsewhere, seems to be answering the 'What is X?' question before asking whether X is Y. For example, if we look now to Book I, do we not find that the question 'What is justice?' is the first question of *Republic* Book I (331c, 331d-336a, 336b-343a)? And is it not *Thrasymachus* who changes the question to

‘Does justice or injustice make us happier?’ at 343bff, esp. 347e-354a? And does Socrates not draw Thrasymachus *back* to the original question, in insisting that we cannot answer the question whether justice will make us happier before we answer the question ‘What is justice?’ (354b1-c3, as well as 350e11-351a2)? As for Books II-IV, is it not apparent also that almost the whole of Socrates’ reply to the question about greater happiness in these books is taken up with answering the question ‘What is Justice?’, then replying in very short order indeed on the question whether it is, in itself, more profitable? (368e-444c, compared to 444c-445b: passages discussed below.) Are we not constrained to ask:

Is Plato not the philosopher who insists – to all but unanimous dissent from modern analytical philosophers – that before we can know whether, say, justice is more profitable, or is teachable, or whatever, we must first know what justice *is*? Is this not the insistence labeled by moderns who claim to know better, ‘the Socratic fallacy’? Is the Plato who commits the alleged ‘Socratic fallacy’ likely to have taken his eye off the question ‘What exactly is it we are saying does or does not make us happier?’ Yet that is just what charges of changing the subject say Plato is doing. (More on the alleged ‘Socratic fallacy’ in n.57 below).

At any rate, if Plato were right that what Thrasymachus wants to talk about is the real nature of justice, even if that is different from what he thinks it is, then the charge in Sachs’s paper would collapse like a house of cards. What, then, will Sachs say in reply?

Evidently, he must challenge Plato’s claim as to what it is that Thrasymachus wants to talk about. When Thrasymachus makes his challenge, is he talking about (trying to talk about) the real nature of justice, as Plato claims? Does he *intend* to talk about the real nature of justice even if what that is happens to differ from what he conceives it to be? Or is he talking about justice as it is in Thrasymachus’ *conception* of justice (regardless of what the real nature of justice might be), as Sachs is apparently claiming? The issue, it can now be seen, is one of identifying just what Thrasymachus’ challenge *is*.

VI. HOW BEHIND SACHS’S CHARGE LIES THE FREGEAN THEORY OF QUESTIONS

Why does Sachs think that what Thrasymachus’ challenge concerns is truths about Thrasymachus’ conception rather than truths about the real nature of justice? (Or, at any rate, why *would* Sachs have to think this if we asked him to tell us how his argument would run if he had to do without the presupposition that Thrasymachus’ conception of justice does in fact pick out the real nature of justice?) In formulating Sachs’s argument just above, I attributed to Sachs the following position.

[After all,] it’s Thrasymachus’ challenge. So he presumably knows what challenge he is laying down. Is that a challenge about justice as psychological well-adjustment? Prior to Book IV of the *Republic* that conception of justice would hardly so much as have *occurred* to Thrasymachus. So that can’t be what

the challenge is. No, it's surely obvious what the challenge is. It is a challenge about justice on the vulgar conception of it.

At any rate, I cannot see by what *other* considerations Sachs could possibly argue that, if a dispute were to arise as to whether the challenge concerned the real nature of justice or merely Thrasymachus' conception of it (regardless of what the real nature of justice might be), it would have to be Thrasymachus' conception.

What we have to ask here is: Why is Sachs supposing that Thrasymachus knows what challenge he is laying down? Why does Sachs suppose that Thrasymachus cannot be making some other challenge concerning a real nature of justice that is in large part obscure to him? The answer flows from a certain very impressive argument to which I shall now turn – an argument that is almost ubiquitous within analytical philosophy, and which was given its classic expression by Frege, in the opening paragraphs of his great paper 'Negation' of 1918. The argument concerns what it is to *disagree* with someone, a matter which itself turns on *what a question is*. The argument goes as follows.

Evidently, if Socrates' reply to Thrasymachus is to be relevant, he must be *disagreeing* with Thrasymachus over something. As can be seen from the formulation of the dispute above in terms of (1), (2), and (3), Sachs is saying that it is no good for Plato to have Thrasymachus say, 'Justice on *my* conception of it makes us less happy,' and for Socrates to reply, 'No, justice on *my* conception of it does make us happier.' If that were what happens – it is what Sachs thinks *is* happening in the *Republic* – then Socrates would not be disagreeing with Thrasymachus, and that would render his reply to Thrasymachus irrelevant.

Very well, both sides can surely agree that an adequate interpretation of the challenge-and-reply requires that challenger and replier are disagreeing with each other. But then we need to know what a disagreement is. (It would be no good here, for example, for Sachs to employ a tendentious account of what a disagreement is). So what is a disagreement? Evidently it is giving conflicting answers to the same question. This in turn raises the question 'What is a question?' And this leads the analytical philosopher straight to Frege's theory of questions. For the line of thought that generates Sachs's objection to Plato involves precisely such a Fregean theory of the identity of questions. What, then, is a question? Frege's view of questions – at any rate of Yes or No questions – is that

4 A question asks whether a certain thought is true or false.

Now we need to know what a *thought* is. Well, thoughts purport to describe facts. So is the fact of the matter about whether the thought is true (the truth value), *part* of the thought in question? It cannot be, says Frege, since

5 if the fact of the matter (about whether the thought is true) is *part* of the thought, we will not know what question is being asked until we know

what the fact of the matter is – i.e., until we know whether or not the thought is true, i.e., until we know what the *answer* to the question is.

But we *do* think (and are surely *right* to think), the argument goes, that, in general,

- 6** we know what question is being asked before we know what the answer is.

Consider only Frege's example:

Is $(21/20)^{100}$ greater than $^{10}\sqrt{(10^{21})}$?

There are not so many who think they know what question is being asked here who will know what the answer in fact is.

More generally, if, contra (6), we do not know what question we are trying to answer before we know the answer, how can rational inquiry, or indeed inquiry of any kind get off the ground? How do we know where to look for the answer to a question if we don't know what the question is till we know the answer?²⁹ And, turning to the matter of disagreement, how can the two parties to the disagreement agree upon methods of settling the question at issue if they cannot agree on what the question is? For if (6) is not true, they cannot agree even upon what the question is, since each must read into their account of what the question is their own (conflicting) answers.

For good measure here, Frege adds that if the fact of the matter about the thought *were* a part of the thought, then a false thought would be just as contradictory as a 'thought which has no being.' Applying this point to disagreement, the person who is in the wrong in a disagreement would always be contradicting himself or herself. And the person who is in the right would be uttering 'the merest of tautologies' (as people say who think of identities of the sort ' $a = a$ ' as saying nothing). Take the present example. Suppose it were actually true that justice is psychological well-adjustment, and also true that the just are happier than the unjust. Then, on the supposition that the fact of the matter about the thought is a part of that thought, Thrasymachus' answer to the question whether the just are happier than the unjust would be:

- 7a** This justice which makes us happier does not make us happier.

On the other hand, if it were true that justice actually makes us *unhappier*, Socrates' answer would be this:

- 7b** This justice which does not make us happier makes us happier.

We would have self-contradiction in either case. By the same token – so the argument goes – the correct answers would in either case be the merest tautologies, for example,

7a* This justice which makes us happier makes us happier, and

7b* This justice which does not makes us happier does not make us happier.

Such consequences Frege takes merely to strengthen his conclusion.

Thus, if we are to be able intelligibly to ask the question ‘Does justice make us happier?’ without yet knowing the answer – or if we are to be able intelligibly to disagree over what the answer is to that question without either of us contradicting ourselves – it cannot be the case that the truth of the matter about justice is part of the question.

To avoid these consequences, logicians like Frege commit themselves to the view that the question expressed by the sentence ‘Does justice make us happier?’ is not about the fact of the matter about the thought – the truth value determined by the thought. Rather the question can only be asking about the thought independently of the fact of the matter – the thought independently of the truth-value determined by the thought when the thought is applied to the world. On this view of what a thought is, it is then very plausible for Frege to identify the thought with the *meaning* of the sentence (cf. n. 32 below), that is, with the *proposition*. (What the *proposition* is, in Frege’s theory, is *independent* of its truth-value – it is designed to be the same proposition whether it is true or false, the same proposition in possible worlds in which it is true and in possible worlds in which *it* – that same proposition – is false).

Correspondingly, we may say that the question ‘Does justice make us happier?’ is asking not about the actual *reference* of the word ‘justice,’ but about the *meaning* of ‘justice.’ (For the reference of a sentence is constituted by the reference of its parts, and the meaning of a sentence by the *meaning* of its parts. The thought, therefore, being the meaning of the sentence, does not contain the reference of the subject-term; so, if the identity of the question ‘Does justice make us happier?’ is given by the meaning of that sentence, the question is not a question about the *reference* of ‘justice,’ but about the *meaning* of ‘justice’).

Elaborating Frege’s point, therefore, in terms of subject-terms instead of in terms of sentences, the question expressed by the sentence ‘Does justice make us happier?’ is not about what is true of the actual *reference* of ‘justice.’ For, once more, if it were, the question would be about, say, that justice which makes us happier, asking of it trivially whether it makes us happier. Instead, it is about the *sense* or *meaning* of ‘justice.’ The affirmer and the denier are to be taken as agreeing that what they are discussing is the *meaning* of ‘justice.’ (They *cannot* agree on the reference of ‘justice’).

It is true that we *could* bring the reference in by amending the Fregean position to affirm that the one holds that the very same meaning determines as reference something which makes us happier, while the other holds – *must* now hold in order to disagree – that that same meaning determines as reference something which does not make us happier.³⁰ Looking at things in this way, we get that what Thrasymachus’ challenge says is that

8m the meaning of ‘justice’ is such that whatever reference is determined by it makes people less happy,

and Socrates must say in reply – if he wishes relevantly to disagree with Thrasymachus, as he plainly does – that

9m the meaning of ‘justice’ is such that whatever reference is determined by it – the same meaning as that mentioned in (8m) – makes people *more* happy.

This account of the disagreement puts us in a position to see that Sachs’s conceptions function here in just the same way as Frege’s meanings or senses. Re-phrasing in terms parallel to those in (8m) and (9m), we would have, on Sachs’s view, that Thrasymachus thinks that

8c the vulgar conception of justice is such that whatever reference is determined by it makes people less happy,

so that, according to Sachs, Socrates must say in reply – if he wishes relevantly to disagree with Thrasymachus, as he plainly does – that

9c the vulgar conception of justice is such that whatever reference is determined by it makes people *more* happy.

The whole idea that lies behind Sachs’s requiring Socrates to use Thrasymachus’ conception of justice if he is to give a relevant reply, is so that we can have a framing of the question which is neutral on such matters as who is right about what justice *is* – on what the reference of ‘justice’ is, and hence also on whether or not that reference makes us happier. (Of course ‘reference’ here only stands for *whatever reference is determined by the meaning*. The assumption is thus being made that meaning determines reference. It does not stand, except in coincidence, for the real reference as I have been speaking of it). Once more, if Socrates were right that *what justice is*, is psychological well-adjustment, and that those who are psychologically well-adjusted are always happier than the ill-adjusted, then if the question were, contrary to Sachs’s view, about what justice really is (the reference of ‘justice’), it would be asking whether that psychological well-adjustment that always brings more happiness than ill-adjustment brings, always brings more happiness than ill-adjustment brings. So to be fair to Thrasymachus, the argument runs, we must turn the question from a question about what justice really is into a question about Thrasymachus’ conception.³¹

Notice one consequence of Sachs’s position. In any debate about any subject, the challenger gets to set the rules of the debate that follows! The challenger gets to use *his* (or *her*) conception of the subject-matter of the disagreement. And, by the way, it doesn’t matter at all whether or not this conception tramples on the truth of the matter that the

debate is over. As Frege's theory of questions makes clear, how we determine *what the challenge is*, is not by what the truth of the matter is over what is being disagreed about, but by the challenger's *conception* of what is being disagreed about. Even if that *conception* does not determine as reference the real nature of whatever it is that the disagreement is actually about!

May we not suspect that something has gone wrong here?

As (6) above indicates, from the Frege/Sachs point of view, the crucial thing here for the theory of inquiry – since we need to know what the question is before we know what the answer is – is that if we are to have a theory of what it is for two people coherently to disagree about something, we need to specify the inquiry or what the disagreement is about *neutrally*, that is, independently of what the truth of the matter is on the question at issue. (Hence, it becomes natural for many philosophers to think of themselves as examining issues from a neutral standpoint.) Now, for those who believe in meanings, a meaning of 'justice,' like a conception of justice, is something which is designed to have its identity quite independently of what the truth actually is about the reference (if any) determined by that meaning. Meanings are what I call *phenomenological entities*, entities we can have knowledge about without knowing what reference (if any) they determine in the real world. If Thrasymachus is wrong about what justice is, or, to take another example, about what epilepsy is (suppose he thinks it is *possession by evil spirits*), or if he is wrong to suppose that there is anything at all in the world that corresponds to either sense or meaning, that does not stop us talking about what he *means* by 'justice' or 'epilepsy.' For what he means *does not depend on what the truth is about justice or epilepsy*. That is just the point of (6) above – making what the question is independent of what the answer is. So too for the meaning of the Greek expression for epilepsy, *ho hieros nosos* ('the sacred disease'). The fact that there is no such thing as possession by devils – no reference determined by the meaning of 'the sacred disease' – doesn't stop people assigning a perfectly good meaning to 'the sacred disease' – even though that meaning picks out absolutely nothing in the real world, but only in some fantasy world of those who believe in devils. That is what phenomenological entities are devised *for*. The same applies to the 'thoughts' of which Frege speaks (and the 'propositions' which Frege's modern followers speak of), and that have been a staple of analytical philosophy throughout the twentieth century. Indeed, these propositions have entered into the very fabric of formal logic as we understand it. They are designed to be the things someone affirms or denies, or even merely considers, independently of whether they are true or false.³²

Finally, it should be evident by now that Sachs's conceptions are precisely designed to be such phenomenological entities. What Sachs *needs* conceptions for is for them to be the conditions imposed upon the question Thrasymachus and Socrates disagree about if Thrasymachus is taken as posing the *question* of the greater profitability of injustice, and offering the negative answer. If Socrates is to *contradict* Thrasymachus, i.e., to be offering an *answer* to Thrasymachus' question that contradicts Thrasymachus' answer, and not simply to be changing the subject to another question altogether, and if

Socrates is to employ a fair and neutral characterization of what that question is, the argument goes, then the question must be frameable independently of what the answer is. So to be fair to Thrasymachus, we have to start, neutrally, with *his conception of justice*. And of course, that conception of justice must itself be frameable independently of what the truth is about justice. So in insisting that Thrasymachus' conception of justice is the one Socrates must employ in responding to Thrasymachus, Sachs must be supposing that *what Thrasymachus' conception is* is independent of the truth about justice. That is, Sachs must be taking the subject of the question to be what I have been calling a neutral, *phenomenological entity*. So the disagreement between Socrates and Thrasymachus may be posed as a matter of Socrates disagreeing about *justice as it is according to Thrasymachus' conception* (as if there *were* such a thing as what that conception determines) without so much as raising the question – as Sachs never does – as to what justice is, or even, indeed, raising the question whether there *is* such a thing as justice.

Hence it is that Sachs thinks that Socrates and Thrasymachus can only agree about what they are disagreeing about, if one of them thinks one thing about the Thrasymachean (or vulgar) conception, and the other denies it. Hence, once more, the issue can only be between

8c the vulgar conception of justice is such that those who are just according to it are less happy,

and

9c the vulgar conception of justice is such that those who are just according to it are *more* happy.

The affirmer holds that the conception in question picks out (as reference, we might say) what makes people happier, while the denier holds (or *should* hold, if, unlike Socrates according to Sachs, he knows what he is about) that the conception in question does not pick out (as reference) what makes people happier. Thus, that one and the same thing they must agree they are discussing if they are even to have a chance at disagreeing, is not the reference (that justice which in fact makes us happiest, if so it be), but Thrasymachus' *conception* of justice.

Such is the basis of Sachs's position, that (a) enables Sachs to ignore the question what the correct answer is to the question 'What is justice?'; that (b) allows Thrasymachus to set the rules for the debate that follows; and (c) that forces Socrates either to use the conception (or meaning) Thrasymachus is working with, or else have his reply be irrelevant to Thrasymachus' challenge.³³

VII. A PLATONIC REPLY TO THE FREGEAN THEORY OF QUESTIONS: ONE *DOESN'T* NEED TO KNOW WHAT THE QUESTION IS BEFORE ONE CAN LOOK FOR THE ANSWER

For Sachs, then, when Thrasymachus says justice doesn't always make us happiest, and Socrates denies what Thrasymachus says, it will have to be Thrasymachus'

conception of justice that they are disagreeing about. No one who accepts (4)-(6) above, namely (to repeat),

- 4 A question asks whether a certain thought is true or false;
- 5 if the fact of the matter (about whether the thought is true) is *part* of the thought, we will not know what question is being asked until we know what the fact of the matter is – i.e., until we know whether or not the thought is true, i.e., until we know what the *answer* to the question is;
- 6 we know what question is being asked before we know what the answer is;

is in a position to deny that the question at issue must be independent of the question what justice *is*. And anyone who grants this independence will have to grant Sachs's contention. So if I wish to insist that the disagreement between Thrasymachus and Socrates concerns not Thrasymachus' conception of justice, but what justice *is* (whether or not Thrasymachus' conception succeeds in picking it out), I will have to dissent from one of (4) - (6).

Now not only do I dissent from (6), that is, the claim that we have to know what question we are asking before we know (or can even look for) the answer; I also deny a key supposition of the Sachs/Frege argument to the effect that the *reference* of 'justice' in Thrasymachus' claim that 'justice doesn't always make us happier' is determined by its meaning (or by the conception employed in the claim). For I do not in general grant that meaning determines reference. As I see it, when Thrasymachus or anyone else makes claims about justice or anything else, their intention is to pick out whatever it is in the world that the relevant expression ('justice' in this case) *actually stands for* – what justice really *is*, even if what that is is different from what the meaning of that expression determines. I rely here on remarks in sec. II of the paper, on the reply to Santas 2003 in n.28 above, and on the remarks in n. 29 above concerning

- (A) the classic Donnellan case ('the man in the corner drinking a martini');
- (B) the case of Oedipus referring to his 'lawful wedded wife;'
- (C) the case of cutting at *Cratylus* 387a; and
- (D) the case of the Sacred Disease.

What applies to these four examples applies also to 'Justice.' We *all* make mistakes in devising expressions (not to say conceptions) for referring to things in the world. Knowing that, it is our intention that if we get things wrong we be understood to refer to the actual thing, not what (if anything) our mistaken description picks out. (If Oedipus uses the expression 'My lawfully wedded wife,' where the meaning of that expression

cannot determine Jocasta as referent, since it is not lawful to marry one's mother, he nevertheless wants us to understand in most contexts that it is Jocasta that he is referring to. And indeed he surely is referring to Jocasta in most contexts).

In the same way, I know that the probabilities are all in favor of its being the case that many of my expressions for the things I refer to are false of the things in question. When I make such mistakes, what I want *you* to do, is to *override* such errors – in favor of how it actually is with the thing referred to. (Even what Oedipus *thinks* he is referring to isn't just 'my lawful wedded wife': it is my *lawful wedded wife who is identical with Jocasta as she is in herself even if she is different from what I think she is*). Thus, contrary to the 'meaning determines reference' theory of how an expression like 'the sacred disease' refers, I maintain the position in sec. II above: that when pre-Hippocratean Greeks used the expression 'The sacred disease,' what they wanted was, in general (e.g., when sending for a doctor, or indeed to anyone who might help), that they should be taken as referring to whatever the actual disease *is* (a neurological, not a god-sent disorder, if so it be), not to what the meaning of the words 'the sacred disease' determine (certainly if, as it happens, the meaning of those words determines nothing at all!). As I put it earlier, such pre-Hippocratic Greeks are to be taken as issuing a blank check, into which hearers are invited to insert whatever the actual disease is.³⁴

I say the same about Thrasymachus. When he uses the word 'justice,' he wants to refer to what justice really *is*, even if the meanings of his words do not succeed in picking it out. It is whatever the real thing, justice, is – something pitiable if he is right, something beneficial if Socrates is right – that he wants to disagree with Socrates over. And that is what Socrates wants to disagree with Thrasymachus over. (I have already offered some defense of this view in nn.28, 30 above).

Someone still unconvinced might at this point suggest the following. 'Look, when Thrasymachus says that

10a The unjust person is happier than the just person,

all he has in mind is that

10b The person who, even in situations in which he can get away with it, keeps promises, doesn't steal, doesn't take advantage of others, etc., is a sap;

Thrasymachus doesn't *care* what justice is.'

As we see, here the question of what justice *is* cannot arise – a point that seems reinforced by the absence of any word for justice or injustice in the analysans (10b). To such a position, I would reply that what Thrasymachus has in mind in (10a) is not (10b), but rather

- 10c** The person who, even in situations in which he can get away with it, keeps promises, doesn't steal, doesn't take advantage of others, etc. – which is all there *is* to justice – is a sap.

The analysans (10c) has the considerable advantage that the word 'just' does occur in it, so that it *is* now suggestive of why Thrasymachus feels he can use the word in the analysandum (10a). But then it is clear that (10c) *does* commit Thrasymachus to a position on what justice is, a position which, if incorrect, exposes Thrasymachus to refutation on just those grounds.³⁵

I would of course freely grant that if such a refutation is to be not merely refutation on a technicality, it had better be the case that a *correct* account of what justice is will also have as one of its near or remote consequences the position that, in general, people who, on grounds of what justice requires, keep their promises, etc. are *not* saps. I shall discuss that point below in secs. VIII-IX.

But I have yet to face the following difficulty: If I wish to say that Socrates and Thrasymachus want to be taken as disagreeing over the real nature of justice – whatever that may be, pitiable or beneficial – then I must admit that

- 6*** at least one of Socrates and Thrasymachus does not know what the question is to which they are giving different answers.

[Thus, as I have remarked above, it is (6) of (4)-(6) that I give up – though giving up (6) will require giving up the Fregean account of what the thought is.] But I do not think this admission is fatal to my position. Still, I need to say how I can maintain that Thrasymachus (for it is he) doesn't understand what question he is asking. How can we be giving an answer to a question if we don't know what the question is that we are answering?

My reply is: Why *shouldn't* it be the case that we do not understand exactly what we are asking until we come upon the answer? Why shouldn't it be the case that when scientists asks what an acid is, they only have a half-baked idea of just what they are asking? *Provided*, of course, that something like the following occurs: First, that

they gain a better idea of what they are asking once they come upon a new and promising answer to their question – say, that it is something soluble in water, and that it is of a sour taste in water solution;

then, second, that

with this more promising idea of the question they were first asking when they earlier asked what an acid is, they are enabled to get a better answer to a second effort to ask their question – say, that it is stuff that turns blue litmus paper red;

then, third, that

with this better idea of what these acids are of which they were asking in their second effort, and so a better understanding of the question they were asking, they get an even better answer – say, in terms of acids being proton-donors;

and so on, indefinitely, with ever-improving ideas of what the question is that they are asking and about the nature of thing they are asking about. Why should this not be what it is like to ask questions and to disagree on answers?³⁶ (I believe, indeed, that it is just what happens in Socratic dialectic). Why do we need that we know *with certainty and determinately* what the question is, before we can look for the answer?

These considerations, I submit, block the Sachs/Frege position from forcing us off of the natural answer to the question ‘What are Socrates and Thrasymachus disagreeing about?’ The Sachs/Frege position asks us to have Socrates and Thrasymachus disagreeing about what is true of *Thrasymachus’ conception of justice* – defending that request by adducing (4)-(6). I reply that the considerations in the preceding paragraph make it perfectly intuitive to deny (6). Once it has been shown that we do not need to grant (6), there is no longer any reason not to give (what is surely) the natural answer to the question ‘What are Socrates and Thrasymachus disagreeing about?’ *It is the real nature of justice that they are disagreeing about.* Not only that, it is that real nature of justice which each would *agree* they are disagreeing about – each granting that such shortcomings as there are in their ways of referring to it, or in their *conceptions* of it, are to be overridden by what the truth actually is about justice (whether it turn out to be a pitiable or a beneficial attribute). Such is the point I am making in using (10c) four paragraphs back, as well as in using (c) in the reply to Santas in n.28 above.

So while Sachs would apparently suppose he can make out a charge of irrelevance without raising the question whether Socrates is right or not in his well-adjustment account of what justice is, I say this is wrong. We *do* need to know what justice is if we are to know what we are asking when we ask whether or not it is the best policy. But we can perfectly well *ask* whether justice makes us happier (or what justice is) without knowing what justice is or even what exactly our question is. If Socrates is *right* in his account of what justice is – that it is a certain kind of well-adjustment – and if he is right in his claim that the well-adjusted (by this account) are happier than the ill-adjusted, then he gets the desired conclusion. There is no fallacy of equivocation here, and no fallacy of irrelevance. If there is anything to the philosophical assumptions I have been defending here, Sachs’s objection, as an objection to *Plato*, will turn out to be entirely without merit.³⁷

VIII. THE BROADER QUESTION OF CHANGING THE SUBJECT: SKETCH OF A REPLY

Suppose now that I have defeated any objection to *Republic* II-IV of the sort Sachs makes – of the sort that argues that what Socrates *means* by ‘justice’ or *dikaiousunē* is too different from what Thrasymachus, Glaucon, or Adeimantus mean for Socrates’

replies to be in the slightest relevant to the claims of Thrasymachus, Glaucon, and Adeimantus which he claims to be addressing. With such objections out of the way, there still remains the broader suspicion with which we started in sec. III above, that says simply that Socrates in *Republic* II-IV has the wrong reference for ‘justice’ or *dikaïosunê*. What of the objection that says – leaving aside all talk of meanings or conceptions – that the reference of ‘justice’ or *dikaïosunê* simply *could not* be Plato’s state of psychological well-adjustment, since it leaves out too many fundamental ingredients of anything we could suppose justice to be, e.g., fairness and morality – a *moral* concern for the balancing of one’s own good against the good of others. So far, it will be quite correctly urged, I have not made so much as a *prima facie* case that ‘justice’ or *dikaïosunê* *could* refer to a state of psychological well-adjustment – let alone, returning to the more Socratic account in Book I (cf. the last three paras. of sec. IV above), to an expertise pure and simple at getting one’s own happiness.³⁸

In the present section, I can offer no more than a sketch of what such a *prima facie* case would look like. I shall in fact make the case not for the Platonic account of justice (as a state of psychological well-adjustment) but for the somewhat different account of justice (as an expertise pure and simple at getting one’s own happiness) that we find in Socrates. For I take it that, since the broader objection would apply to Socratic expertise in exactly the same way it would to Platonic psychological well-adjustment, showing how the Socratic account can meet the broader objection will suffice for purposes of showing that the Platonic account could also meet it (merely by being at least as good an account of justice as the Socratic account is).

My strategy of responding not on behalf of Plato’s psychological well-adjustment but on behalf of Socratic expertise has three advantages, as I see it. First, I believe that if *either* the Socratic or Platonic account of justice is correct, it is the Socratic account. (In fact, I believe the Socratic account does give the correct account of what justice is. I think this because of the superiority I assign to the Socratic psychology of action over Plato’s parts-of-the-soul psychology of action, and indeed – to chance my arm here a little – over all other psychologies of action). Second, if we see how the Socratic account responds to the broader objection, we should be able also to see how this extraordinarily inward-looking and self-regarding Platonic account develops *from* the Socratic account. This is so, I think, because, as I believe, the Platonic account develops from the Socratic account solely by a change in the Platonic psychology of action – from ‘Socratic Intellectualism’ to a ‘parts of the soul’ doctrine. This parts of the soul doctrine, which I shall call Plato’s ‘Limited Psychological Rationalism,’ involves an explicit³⁹ rejection of Socratic Intellectualism – the Socratic psychology of action. As I have just suggested, my reason for thinking the Socratic account more correct than the Platonic account resides solely in the fact that I regard Socratic Intellectualism as a more correct psychology of action than Plato’s Limited Psychological Rationalism (the parts of the soul doctrine). Hence, I believe that seeing how the Platonic account develops from the Socratic account merely by this change in the psychology of action should make it clearer how Plato would have made *his* *prima facie* case for the reference of ‘justice’ being a state of psychological well-adjustment. Third, and closely related to the preceding point, if we see how the

Platonic account develops from the Socratic account, we will also be able to understand the shocking feature of the Platonic account which I made much of in sec. IV above: that the Platonic account entirely abandons the idea of characterizing the justice of an individual in terms of the supposed justice of a citizen of the just city (= that citizen's standing in the appropriate external relations to other citizens of the just city). A treatise primarily concerned with political arrangements, of the sort we find in the *Laws* and even in the sort of treatise that *Timaeus* 17c-19c purports to be abstracting, might be expected to show a primary interest in the external relations of citizens to each other. But the primary concern of the *Republic* is with whether the just individual is happier than the unjust individual. Its concern with details of political arrangements is at best incidental and in digressions. (See n. 1 above).

To return to the astonishing suggestion that justice is a certain sort of psychological well-adjustment, a very great deal of what Plato says is opaque to modern thinkers. One way of looking at what I am suggesting in this section is this: The reason people have had such a hard time seeing even a prima facie case for justice being psychological well-adjustment is that *they have not been seeing it as emerging from the Socratic account of justice*.

In my sketch, therefore, I need to say enough to bring out both how the Socratic account of justice differs from the Platonic account in its psychology of action, and also how, psychology of action aside, the Platonic picture of justice does not *further* differ from the Socratic account, and does indeed give us an account of the virtue of justice that, like the Socratic account, is egoistic in the following way (though only in the following way: cf. n.16 above): the just or good person will be the person good at maximizing his or her own happiness (together with whatever this maximization brings with it by way of the good of others).

Because this is just a sketch of the Socratic account, many details will be given only the most incomplete justification, and some will be given none at all, given that even the most incomplete justification will take us too far afield. Fuller justifications for some of these features have been given, or will be given, elsewhere.⁴⁰ Without this supporting material, it may be less than clear why I myself find this picture so convincing. That is the price one pays for trying to deal with large themes in a relatively short space. But the aim here is merely to sketch the sort of way in which the prima facie case can be made for both Socrates and Plato thinking that their account of the reference of 'justice' could perfectly well be what justice *is* – and so be what even Thrasymachus, Glaucon, and Adeimantus are asking about, even though this justice is not at all *moral*, but entirely egoistic (as per the account in the preceding paragraph).

Here, then, very briefly, are the main ideas: First, Socrates is a 'psychological egoist' of the following sort: He holds that

PsE every action whatever [at any rate any action of the sort *we* tend to call 'voluntary'] is determined by precisely two things (a) a certain

fundamental desire every agent, x , has for whatever action now will bring x the maximum real good (real happiness) available to x over a complete life, starting from where x is now; and (b) x 's belief at the time as to which particular action that is, that belief being produced by a certain pathway through x 's entire web of belief – a pathway running from that fundamental desire to the particular action the agent does. The action the belief identifies as the action which will bring x the maximum real good (etc.) is then substituted into the fundamental desire to generate the action.

The Platonic psychology of action, as I construe it, differs precisely here. For Plato, on the one hand, *some* actions – those proceeding from the rational part without interference from the two irrational parts of the soul – are of just the sort that *all* actions are according to Socrates. They proceed from that single fundamental desire in the rational part of the soul for whatever action now will maximize the real good of the entire individual together with a belief as to what that action is. On the other hand, in Plato, actions may also result from desires of the spirited part only (when they occur together with some beliefs), or from desires of the appetitive part only (when they occur together with some beliefs). On my reading, neither of the latter two sorts of desires are maximizing desires for good (n.39 above). The result is that we get in Plato, what we never get in Socrates, that, due to irrational desires, one may act ('voluntarily', as Aristotle and we would say) contrary to one's desire for whatever action now will maximize one's own real good. So, while Socrates is a psychological egoist, Plato is not. (I will not here attempt to assess the proposal of a further development in Platonic psychology of action in Books VIII-IX, where another pathology may seem to be involved – where reason appears to be non-intellectually corrupted in its beliefs as to what is good by certain habits of indulging irrational desires. If this proposal is correct, then Plato gives us in Book IV, Aristotelian *akrasia*; in Books VIII-IX, something like Aristotelian *akolasia*: cf. n. 39 above).

Returning to the Socratic schema (PsE), not only is every action determined by the fundamental desire and the pathway taken through the web of belief. It is also the case, in my view, that

- Det** Both this desire for the real good and this belief are determined:
 (a) the desire is determined because this desire belongs to all of us by something like what we would call a 'law of nature'; and (b) the belief is determined because (as I suppose – there is nothing explicit on this in the dialogues) it is determined by the agent's prior intellectual state together with whatever perceptions, thoughts, or dialectical conversations have occurred to the agent since then.

The idea behind the *belief* being determined is roughly that you don't get to choose what you believe true, any more than that you get to choose whether or not to desire the real good. (One does not get to choose *now* one's prior intellectual state. But the world, together with that prior intellectual state, is responsible for the perceptual states that occur during the time that intervenes between that prior state and the present, as also for

intervening encounters with others with whom one interacts in discussion. And one does not appear any more to be responsible for what new thoughts come to appear true to one in that intervening period. What appears true, all things considered, *forces itself* on us. What we believe is just as much determined as what we desire).

The presence of such a deterministic account of the psychology of action seems to me confirmed by the way it makes intelligible the idea that virtue can only be knowledge and vice only ignorance. For in Socratic psychology of action, good people do not differ from bad people in the fundamental desire that generates action (*Meno* 78b4-6). We may infer, then, at any rate on this reading of the Socratic psychology of action, that they can differ only in their beliefs. Differ how? One can only suppose in the fact that the good are better *at* (cf. the end of sec. II above, as well as nn.16, 24, and 44) what they do in pursuit of their own maximal happiness by virtue of having truer beliefs (or indeed, if this ever happens amongst humans, knowledge).

So I claim at least the following merit for my exegetical assumptions above: that they explain perhaps better than any other account the most well-recognized and central of all Socratic doctrines: that virtue is knowledge and vice ignorance. They also explain other characteristic views, such as that

knowledge [presumably together with what it is causally inseparable from, given the nature of our fundamental desire for the good, that being desire for the real good] is the only thing good in itself,⁴¹

and that no one errs willingly [at getting their own real good], i.e.,

all error is due to ignorance.⁴²

Now, given this determinism, I conclude immediately that all theories of human goodness in terms of *moral* goodness are ruled out. For I take it that

MORAL Morality requires *could have done otherwise* for voluntary actions (*ought implies can*),

and by Psychological Egoism (PsE) and the deterministic principle (Det), no one ever could have done otherwise than they actually did.

Here I make two assumptions: First, I assume that compatibilism cannot deliver a *morally* relevant ‘could have done otherwise.’ Of course a *politically* or *administratively* relevant ‘could have done otherwise’ is very often endorsed by able philosophers – for example, Sidgwick 1907, 71-2 – who propose not to allow their view of civilized social arrangements and political order to be thrown off by worries about determinism. The result is, however, that they have to swallow certain beliefs about the efficacy of punishment as a device of reform and deterrence that, so far as I can see, only the most

administratively or politically desperate could buy into. So I do not myself consider this option taken by determinist politicians and administrators to be at all viable.

I also assume that we must include under (MORAL) all three of Sidgwick's methods of ethics. This will be obvious for all so-called 'deontological' theories that fall under the method of Intuitionism. But it also holds for the so-called 'teleological' ethical theories that fall under Sidgwick's Egoistic and Utilitarian methods of ethics. For both of those two methods of ethics, in spite of the fact that 'good' is a fundamental notion of which they make use, require a categorical imperative: You *ought* to promote your own good, you *ought* to promote the Greatest Happiness of the Greatest Number.⁴³ Hence all three of Sidgwick's methods of ethics require *could have done otherwise* and so fall under (MORAL).

So, now, given the rejection of (MORAL) and the claim four paragraphs back that good people differ from bad people solely by the good having truer beliefs, we get the following important consequence for ethics:

SE The only remaining viable theory of human goodness is that which we find in 'Socratic Ethics' – a purely factual theory that modern moral philosophers might refuse to call an ethical theory at all, since there is nothing of 'morality' to it, and nothing of 'moral values' to it. Socratic Ethics comprises both (a) the psychological egoism formulated above as (PsE) and which I also call 'Socratic Intellectualism,' and (b) the view that, as a purely factual matter, the good person is the person good at maximizing that person's own good.

(I leave aside here the many forms of relativism, since I do not regard them as relevant to the context of interpreting Socrates or Plato).

Notice that we need to be a little careful about calling Socratic Ethics, as per (SE) a form of egoism, or even for saying that part (b) of (SE) – that is, the thesis that as a purely factual matter, the good person is the person good at maximizing his or her own good or happiness – is a form of 'Ethical Egoism.' Care is necessary for the simple reason that most of those philosophers who use the expression 'ethical egoism' think of ethical egoism in terms of a *moral obligation* to pursue one's own good, or in terms of certain *moral values* or normative prescriptions favoring one's own good (as in Sidgwick's egoist 'method of ethics'). Those conceptions do not apply to Socratic Egoism as I am presenting it here. For Socrates, it is just as much a matter of (psychological) fact that (b) the good person is the person good at maximizing his or her own good, as it is that (a) every action is determined by the agent's desire to maximize his or her own good and his or her beliefs. There is no contrast between (a) and (b) in terms of *ethical* vs. *psychological*.⁴⁴

It is true that, at first blush, making Socratic Egoism the only viable theory of human goodness seems to have the obnoxious consequence that

Obn If it is the part of a good person to gain his or her own good, then it may well be the part of the good person to gain his or her own good at the expense of others.

If the possibility envisaged in the consequent of this obnoxious conditional should be realized by its actually being the case that a good person gains his or her own good at the expense of others, then the best we could get out of justice is what Thrasymachus has on offer – either the exploiting of others with a view to one’s own benefit, or (thanks to the improvement Socrates forces on this account) the *science of exploiting others for one’s own benefit*. (Socrates would become a Thrasymachean!) Such a science of exploiting others for your own benefit surely cannot be the reference of ‘justice.’

The difficulty here may seem to suggest that one should reject this interpretation of the Socratic parts of the Platonic corpus in terms of what I have called ‘Socratic Ethics,’ and seek to pursue instead the line of thought in some scholars to the effect that however egoistic Socrates may appear to be (both in his psychology of action and in his theory of human goodness), when he comes to the *Apology* and the *Crito*, Socrates goes well beyond this Socratic Egoism (plainly enough present in other early dialogues) to the following moral position:

11m It is morally wrong ever to harm others. (We have a *moral obligation* to do no harm to others).⁴⁵

If these scholars are right about the *Apology* and *Crito*, then my account (SE) above of Socratic Ethics cannot be right.

I can here only sketch the briefest of replies to the objection that at least the *Apology* and *Crito* show Socrates committed to the existence of a moral good over and above one’s own self-interest. My claim is this: The position that Plato actually espouses in the *Apology* and *Crito* is not what we have in the claim that it is morally wrong ever to harm others, even in retaliation, but rather the following purely factual claim:

HARM It is never in your interest to harm others. (It is a *fact* that to harm others is to bring harm to yourself).

If this is right, then seeking your own maximum happiness can never have the obnoxious consequence threatened by our conditional above:

Obn If it is the part of a good person to gain his or her own good, then it may well be the part of the good person to gain his or her own good at the expense of others.

For if, by (HARM), it is never in your interest to harm others, then we will never be able to derive from the truth of the antecedent the Thrasymachean conclusion that anyone ever

will gain his or her own good at the expense of others.⁴⁶ This, of course, in spite of the fact that the theory of desire in the Socratic psychology of action is entirely egoistic. This shows, if such showing was needed, that *Egoistic* is not the same as *inclined to seek one's own good at the expense of others*. (*Self-interest* is not the same as *selfishness*.) For on the hypothesis (HARM), a knowing egoist will never attempt to gain his or her good at the expense of others – since there is *no such thing as* gaining your own good at the expense of others. The Socratic science of advantage is not the Thrasymachean science of gaining one's own advantage at the expense of others.⁴⁷

Now it is true that my interpretation depends upon the correctness of my (HARM). I freely admit that if I am wrong in this reading of the *Apology* and *Crito*, then the rest of my argument in this section of the paper will fail. But for better or worse, my sketch *is* made on the basis of the assumptions outlined above, including (HARM).⁴⁸

At any rate, I have now presented the 'Socratic Ethics' which I believe is derived from Socratic Psychological Egoism. I am the more inclined to think there is something to my interpretation since I myself strongly endorse the Socratic psychology of action. Given this, and my embracing both determinism and the claim (HARM), I am also strongly inclined to endorse the resulting Socratic Ethics, with all its radical consequences.

But what matters for present purposes is how I am to find in this account which Socrates gets from his psychology (which is also a theory of human goodness) a *prima facie* case that the account of justice that shows up within this Socratic Ethics I have been describing avoids the charge of giving us entirely the wrong reference for the word 'justice.' To this I turn immediately.

With the above assumptions (PsE), (Det), (SE), and (HARM) in hand, I can lay out what I take the Socratic theory of justice to say. My view of the Socratic account of justice – itself a product of the Socratic psychology of action – I extrapolate from the doctrine of the Unity of Virtue that we find in the *Protagoras*:

- UV** 'Courage,' 'temperance,' 'justice,' 'piety' and 'wisdom,' whatever those words may *mean* (and if there *are* any meanings), are five different names of one and the same thing, namely,
- *the science of goods and bads, i.e.,
 - *the science of gaining one's own happiness, i.e.,
 - *the measuring art (an expertise pure and simple), i.e.,
 - *the knowledge of *what is worth trading for what* amongst goods or pleasures at various temporal distances from the present.

Think of the differences in the *words* 'courage,' 'temperance,' and so forth as merely *historical* differences: as with 'Evening Star,' and 'Morning Star.' For historical reasons, we tend to use 'Evening Star' in the evening, 'Morning Star' in the morning. But that is not a difference in the thing we are talking about. For similar *historical* reasons,

*Use ‘temperance’ as the name of this science of happiness in situations where there’s a temptation to think that certain bodily pleasures will maximize one’s happiness,

*Use ‘courage’ as the name of this science of happiness in situations where there’s a temptation to think that fleeing certain dangers will maximize one’s happiness,

and (*here is my extrapolation*)

*Use ‘justice’ as the name of this science of happiness in situations where there’s temptation to think that (*trying to*) *take advantage of others (pleon echein) or to harm others* will maximize one’s happiness.

But now – and this is the whole point of (UV), as also of the claim that virtue is knowledge (the science of goods and bads) – in none of these situations is there a difference in the psychological state we are talking about, only in the historical associations of the word. By the same token, what self-indulgence, cowardice, injustice are, is, again, one and the same thing, namely, ignorance (Vice is ignorance). What ignorance? Ignorance of goods and bads – that ignorance that consists in not seeing that one will *not* be happy giving in to the aforementioned temptations.

Such, I say, is the Socratic account of justice.⁴⁹

It is true that even many devoted readers of Plato will find this account of justice very odd-looking. But is it any odder than the Socratic account of courage as a measuring art for balancing goods and bads (pleasures and pains) against each other in temporal perspective – so that the cowardly and the courageous do not differ in what they aim for (their own over all good or pleasure) but solely in whether or not they possess the measuring art? Or, in the account of temperance as this very same measuring art? But that is surely the view of courage and temperance in the *Protagoras* (see Penner 1996, 1997).

I admit, of course, that while the talk of dealing with the temptations to giving in to beliefs about present fear or present pleasure has textual warrant in the *Protagoras*, we do not have similar warrant for speaking of the temptations to get the better of others by harming them. But given Socrates’ evident interest in the topic of getting the better of others (at any rate in *Republic I* and in the *Gorgias*), we need only note that, while those with knowledge know that getting one’s own good by getting the better of others is something that *simply can’t be done*, those without knowledge, perhaps because of an inadequate grasp of (HARM), may well be tempted to harm others (cf. here *Republic* 366c6-d1). (Notice, incidentally, that all the temptations described above are *intellectual* temptations, tendencies to be taken in intellectually by the appearances of things. The appearances promote the occurrence of beliefs, or promote the salience of some beliefs, that suggest one can do better by cowardly, self-indulgent or unjust deeds, over others

that might point towards action in a different direction. It is only by having *all* of the beliefs in one's web of belief in order, as happens with someone with knowledge, that one is able to defeat these mistaken appearances, or balance these tempting beliefs against others and so see one's way through those tempting appearances.) In speaking of justice as the expertise described in (UV), Socrates is not changing the subject.

I conclude that, given the Socratic position on the unity of virtue, there is a strong *prima facie* case for the view that even though Book I of the *Republic* shows Socrates pushing an account of justice as an expertise or science pure and simple, Plato is nevertheless engaging in no change of subject on us in Book I. The Socrates of Book I may be *wrong* about what justice is; but, if so, *that* is the flaw in his argument, not that he is changing the subject on us.⁵⁰

I now apply these remarks to Plato's psychological well-adjustment. (Briefly, since I hope the general idea is clear by now). The well-adjustment account of justice differs from (UV) solely in its anti-intellectualist (parts of the soul) psychology. There is, I claim, no further difference in the ethics that does not flow from this change of psychology. In Socrates, the just person acts with the knowledge of what is best for that person – a knowledge gained by dialectic alone. In Plato, the just person acts with the knowledge, in his or her *rational part*, of what is best for *the whole person* (all three parts of the soul) – a knowledge gained by dialectic, though it cannot be gained by dialectic alone. *Also* necessary to the gaining of such knowledge is that one start off with the right nature, and also that one undergo an appropriate character-training, largely non-intellectual, directed at all three parts of the soul. In Plato, by contrast with Socrates, the correct nature and training are preconditions for the acquisition of the relevant knowledge. But it is still the case that, for Plato, the good or just person is in fact the person who is good at maximizing his or her own happiness. (This is surely precisely what one must suppose if the whole point of the *Republic* is that the just are always happier than the unjust, even *without* counting the rewards of reputation).⁵¹ The *Crito*'s argument that the just are happier, by way of the analogy with medicine, is the same as that in the *Republic* – as it is also at *Gorgias* 504d-505b. But in the *Crito*, what is analogous to medicine is dialectic, while in the *Republic*, what is analogous to medicine is – or at any rate absolutely requires – the training of emotions and appetites by non-intellectual means (cf. n.25 above). From my own (Socratic) point of view, what's wrong with the account of justice in the *Republic* is not that it is too internal to the psyche, and not enough concerned with our relations to other people; it's simply that its account of the insides of the psyche is wrong. What is wanted (I have argued elsewhere) is not the Platonic theory of three parts of the soul (or – what, pace Aristotle, is very nearly the same thing – the Aristotelian theory of three kinds of desire leading to what Aristotle calls voluntary action, one rational kind, and two irrational kinds),⁵² but Socratic intellectualism.

Finally, notice that we can now explain *Republic* IV's treatment of justice in the individual purely in terms of relations internal to that individual and, surprisingly for moderns, apparently ignoring the external relations between individual citizens within the

just city. For the Platonic account derives from the Socratic account in (UV) without changes other than those occasioned by a change in the psychology of action – apparently ignoring the good of others. The good of others does enter in, however. But not as a moral matter. Just as a matter of the factual claim in (HARM) that harming others harms you.

IX. SUMMARY, WITH A CONCLUSION ON THE USE OF MEANINGS, PROPOSITIONS, CONCEPTIONS, CONCEPTS AND THE LIKE IN READING PLATO

The preceding section is designed to show how it *could* be the case that Plato, in the *Republic*, really can regard justice as an inner state of psychological well-adjustment that, like health in matters of the well-functioning of the body, serves the agent's self-interest – and that *without* changing the subject on Thrasymachus, Glaucon, and Adeimantus. Let me summarize the preceding section and earlier sections by reviewing several considerations that the preceding sections claim lie behind this conclusion.

First of all, there is the matter of the question Socrates and Thrasymachus (etc.) are to agree they are asking. On the account offered above, the question they are asking is whether the real nature of justice (whatever that may be, and even if what that is is different from how either of them conceives of justice, or from what either of them *thinks* it is) always makes people happier than would the real nature of injustice (whatever *that* may be, and even if *that* is different from how either of them conceives of injustice, or from what either of them thinks it is). On this account of what the question is to which they are giving different answers, it could be the case at the beginning of the inquiry that neither of them knows what justice really is or even what exactly the question is that they are asking. The hope of the dialectical inquiry is, however, that what the very thing is that they are asking about, namely, justice, will emerge more clearly from the inquiry. (Just as happens, in favorable cases, in the history of scientific inquiry: recall the discussion of acids towards the end of sec. VII). All that is necessary is that they are not *so* far off that it prevents the inquiry from proceeding. (It *could* happen that they are too far off, and in that case the inquiry will break down. But in the *Republic*, we are to believe that it does not. But as to how far off one or both interlocutors are, that we will only be able to tell when we get the right answer).

Looked at in this way, there is no difficulty about Socrates and Thrasymachus disagreeing over the same question. A philosopher like Sachs may *persuade* them that they are not discussing the same question, on the grounds that each has a different conception of justice (each has different beliefs about what justice is). But by the present account of questions, what people ask questions about is not whatever their conceptions pick out of the world (which is very often nothing at all, as with Oedipus' lawfully wedded wife – unless one is prepared to abandon the real world for *some other possible world*, some dream world that Oedipus takes himself to be living in, in which Jocasta is not Oedipus' mother!!). The question is about the real nature of whatever it is that is in question, even if what that is is different anything either conception would pick out.

Given this response to Frege's theory of questions, which dispatches the criticism of Sachs, it remains only to ask whether it is at all plausible to suppose that the real nature of justice could be anything so odd as the internal state of psychological well-adjustment that we find in the *Republic*. A similar question might arise, we have seen, about whether it is at all plausible to suppose that the real nature of justice could be anything so odd as the internal state consisting in the knowledge of goods and bads which we find in Socrates. Anyone, we supposed, who would rule out the Platonic answer as implausible right from the start, would surely also suppose that the Socratic answer is implausible right from the start. So if we can show how the Socratic answer might actually be correct, we'll have shown how the Platonic answer could also be correct. Now, on the account offered here, the only differences between the Platonic answer and the Socratic answer are differences that flow from differences in their psychology of action. So if the Socratic psychology of action is correct, then the Platonic theory *would have been correct* had Plato only been right in his psychology of action. So the project would *not* have been ill-conceived in the way it is natural to suppose it is, on the grounds that Platonic justice looks so different from justice as we ordinarily understand it.

Turning then to assessing whether the Socratic account of justice cannot be what justice is, my argument appealed to the Unity of Virtue doctrine. Could that account capture the nature of justice?

To show that it could, I needed (a) my account of the Socratic psychology of action, which I formulated as the version (PsE) of psychological egoism, as well as (b) the assumption that this psychology of action made all actions determined. With these two assumptions in hand, it became natural for us to suppose that Socratic Ethics is the only viable (non-relativist) theory of human goodness. (The moral good, standing in contrast as it does to one's *own* good, requires *could have done otherwise*, which is, for determinists, unavailable). Human goodness can then only be goodness *at* getting that maximum of one's own happiness that is available to one, starting from where one is now. (These assumptions make excellent sense of 'Virtue is knowledge,' 'Vice is ignorance,' 'Being overcome by pleasure is ignorance,' 'No one errs willingly,' and the like).

At this point, there was the worry that this would have the obnoxious consequence that harming others could be a way to become happy. To show I need not face this consequence, I argued that the assumption (c) that the supposed appearance of the moral principle 'You ought never to do injustice back to (or harm back) someone who does injustice to you or harms you' in the *Crito* is really not a moral principle saying we have a moral obligation not to harm others, but rather a factual claim that harming others will harm oneself – so that there could be no such thing as a Thrasymachean science of exploiting others. The obnoxious consequence is avoided.

On all of these assumptions, then, together with the assumption that Socratic Intellectualism is actually *correct* – largely on the grounds that its psychology of action is

correct – we can see how *what justice is*, assuming that there *is* such a thing, could turn out to be precisely this expertise at getting one’s own happiness (which is of course ‘wrapped up in’ the happiness of those around one). With that account of justice in hand, we can also see how a change in Plato’s psychology of action would result in a (mistaken) switch from

justice as this particular expertise at maximizing one’s own happiness

to

justice as that particular state of psychological well-adjustment, part of which includes one’s rational part having the expertise at maximizing one’s overall happiness – an expertise Plato alleges it is impossible to acquire without the appropriate (largely non-rational) training of the two irrational parts.

Notice that this switch involves no change of subject. If (as I suppose) Socratic Intellectualism is true, the justice we all want to talk about and refer to would just *be* (whatever else we might *think* it is) this particular expertise at getting one’s own happiness. If, on the other hand, Plato’s Limited Psychological Rationalism were true, the justice we all want to talk about and refer to would just be (whatever else we might think it is) that particular state of psychological well-adjustment.

The resulting sketch of how the *Republic* might be thinking of justice represents, I realize full well, an approach that is quite different from those of many other interpreters of Plato. In particular, the theory of human goodness which I attribute to Plato is quite different from the moral eudaemonism we find in Irwin 1995 – ‘happiness’ in this ‘eudaemonism’ has morality built into its very meaning (cf. n.14 above) – and the near-Rawlsian ethical theory we find in White 1979. I shall not undertake here to refute these two distinguished and formidable readings of the *Republic* so different from my own. Obviously neither of those accounts could regard the set of assumptions my account makes as acceptable as a whole. But I have not here tried to show those assumptions correct. My aim in the preceding section has been merely to show how the claim can be rendered plausible on certain exegetical and philosophical assumptions.

* * * * *

I want, in conclusion, to return to the question of what various words in the *Republic*, and in our own moral discourse, *mean*. So long as Plato’s readers remain devoted to reading Plato in terms of such phenomenological entities as meanings, propositions, and conceptions à la Sachs, neither the Psychological Egoism I attribute to Socrates nor the theory of human goodness I attribute to Socrates *and* Plato stand much of a chance of establishing themselves as relevant to understanding how Plato should have come to such a view of what justice is. They are too distant from *what we mean* and involve conceptions too different from our own. (Or, as it is sometimes put, they are too distant from ‘our linguistic intuitions’). The search these readers engage in is for a

philosophically neutral platform from which readers can discuss clashes of positions in the dialogues. This is especially true of the *Republic* where the judicious have so long thought it perfectly straightforward that the character Socrates has not responded to the difficulties laid against him by Thrasymachus and the two brothers.

From my own point of view, and the point of view I attribute to Socrates and Plato, the search for such a neutral platform, which leads to the positing of meanings and propositions – generally, to phenomenological entities – is entirely mistaken. What I want to say is: *it makes a difference* to our account of what Socrates is saying or what Plato is saying what the *truth* of the matter is that they are talking about. (It matters even to what the *question* is that we are asking, what the truth is about the entity being asked about). Neutral platforms there are none. Dialectic, between people not fully in the know, is our only hold on what it is we are talking about.

Evidently my account of the Socratic view of justice will establish itself as worth considering only if there is some plausibility to the claims Socrates makes (according to me) about psychology of action, determinism, and the theory of human goodness. Absent any plausibility to these philosophical assumptions, I am saying, we should not be wasting our time on this interpretation – no matter how good our reading of the text otherwise. There is no suspending of our philosophical views, first, to determine what the various interlocutors in the dialogues *mean* by their words; then, second, to raise the question of the relevance of these meanings to what *other* interlocutors mean by *their* words; and then, third, to raise the question whether what they mean is true or not. That way lie accusations of fallacy that flow merely from the (quite un-Socratic and un-Platonic, as well as false) assumption that meaning determines reference.

How do interpreters come to make this error of supposing that we can determine what characters in a dialogue are saying merely by examining what they *mean* – appealing to such phenomenological entities as meanings, propositions, conceptions, and so forth? Partly, it flows from the idea of argument as a discipline philosophers are good at *independently of what the truth is*. (Cf. most treatments of the so-called ‘Socratic elenchus’).⁵³ Here we have a methodological ideal that has been elevated into a metaphysical truth.

Let me recall, here, if I may, how in my 1987, I inveighed against appeals to the principles of [classical] logic and to Fregean principles about the logic of psychological contexts as each metaphysically neutral platforms from which we could assess Platonic metaphysical claims, say, about the Forms. I suggested there that such appeals are, in the first place, *not* metaphysically neutral – given that those principles of logic *themselves* presuppose the existence of entities – logical forms – of exactly the same sort as the Platonic Forms. (Indeed, these entities face, in the paradoxes, much more serious difficulties than the Forms actually face). What is more, retrenchments to avoid the paradoxes simply make the resulting logic *ad hoc* and implausible as an *organon*. Second, I suggested that anyhow to accept the existence of these logical forms and simultaneously to regard the principles of logic as true *independently of all matters of*

fact and real existence – which is after all what metaphysical neutrality would require – is to contradict oneself. How could it be that the principles of logic are true no matter what is true? And that logical forms exist no matter what exists? This, I suggest there, makes the enterprise of using logic as a neutral tool for interpreting Plato a rather more doubtful concern than any of Plato's own arguments for the existence of Platonic Forms.

Returning to the main topic of the present paper, I now offer the following diagnosis of the error I see in appeals to meanings, conceptions, and propositions in declaring that Socrates has or has not succeeded in disagreeing with Thrasymachus, Glaucon, and Adeimantus. I see the error as flowing from the famous quantifier shift, as when one mistakenly infers from 'Every boy loves some girl' that some girl is loved by every boy, that is, from the agreeable scenario in which every boy is sweet-hearted to the disagreeable scenario of the one universal sweetheart. What is abstractly true, I think, about every particular disagreement, is the following:

- 12a** For each thing that may be the truth of the matter the disagreement is about, there is a particular account of what the disagreement is about such that that particular account is the correct account.

By contrast, on the phenomenological (propositional, meaning-based) view, what we have, reversing the quantifiers, is that what is abstractly true about every particular disagreement is this:

- 12b** There is a particular account of what the disagreement is about, such that for each thing that *may* be the truth about the matter the disagreement is about, that particular account is the correct account.⁵⁴

In the first, what the correct account of the disagreement is depends upon what the truth of the matter is that is disagreed about. In the second, there is no such dependence, and a certain absolute (phenomenological) neutrality, given by meanings and propositions. I trust I have conveyed my own lack of faith in the latter.

It is my view, no doubt a controversial one, that it is time interpreters of Ancient Philosophy dropped this idea of appealing to what words mean in trying to settle what the philosophers in question are saying.⁵⁵ The job cannot be done at all except – as generations of our best Plato interpreters have shown (not, alas in their theory, but at least in their practice) – by trying to answer simultaneously the question what Plato is saying (e.g., 'What is Plato saying justice is?') and the question what the truth of the matter is about what Plato is discussing (e.g., 'What is justice really?') So far from its being a 'Socratic fallacy' that to know whether justice makes us happy or is teachable, one must first know what justice is,⁵⁶ Plato is quite right that you cannot know the answer to 'Is x F' without knowing the answer to the question 'What is x .'⁵⁷ None of which stops us from *trying to answer* the question whether justice makes us happier without knowing what justice is. To think otherwise is to fall in with Frege's seductive, but surely false theory of questions.

I conclude that all of Thrasymachus, Glaucon, Adeimantus, and the Socrates of Books II-X intend to refer to the real nature of justice. And so does the Socrates of Book I, and the Socrates of the *Protagoras*. If the psychology of action of the *Protagoras* is correct, then courage *is* what Socrates says it is in that dialogue. But then why should justice not be what Socrates says it is in the *Protagoras*, as elaborated in my (UV) above? And if justice is that, and Plato makes the particular mistake about human psychology that he makes in the parts of the soul doctrine, then what he too intends to refer to, when he speaks of justice as a certain psychological well-adjustment, is also justice as it really is. And – bar that error in his psychology – he is right. He is certainly not refuted by failing to address himself to what Thrasymachus and Glaucon *mean* by the word for justice. Nor is he refuted by the unusual look justice has, once we discover what it is.^{58, 59}

Terry PENNER
University of Wisconsin

BIBLIOGRAPHY

- Annas, J. 1981 *An Introduction to Plato's Republic*. Oxford: Clarendon Press.
- M. Beaney (tr. and ed.) 1997 *The Frege Reader*. London: Blackwell.
- Bobonich, C. 2002 *Plato's Utopia Recast*. Oxford: Clarendon Press.
- Butler, James P. 2002 'Justice and the Fundamental Question of Plato's *Republic*' *Apeiron* 32, 1-17.
- Cooper, J. (1977) "The Psychology of Justice in Plato's *Republic*" *American Philosophical Quarterly*, 14, 151-57.
- Davidson, D. 1967 'Truth and Meaning,' in *Inquiries into Truth and Interpretation* (pp.17-36). Oxford: Clarendon Press.
- Dodds, E.R. 1959 *Plato: Gorgias*. Oxford: Clarendon Press.
- Donnellan, K. 1966 'Reference and Definite Descriptions,' *Philosophical Review* 75, 281-304.
- Evans, G. 1982 John McDowell (ed.) *The Varieties of Reference*. Oxford: Clarendon Press.
- Ferejohn, Michael, 1984 'Socratic thought-experiments and the unity of virtue paradox.' *Phronesis* 29, 105-122.
- Foster, M.B. 1937 'A Mistake of Plato's in the *Republic*,' *Mind* 46, 386-93.
- Frege, G. 1884 *Grundlagen der Arithmetik*, translated in part in Beaney 1997 (pp.84-129).
- 1892, 'On *Sinn* and *Bedeutung*' in Beaney 1997 (pp.151-71).
- 1918 'Negation' in Beaney 1997 (pp.346-61).
- Gadamer, H-G. 1986 *The Idea of the Good in Platonic-Aristotelian Philosophy* (tr. and introd. P. Christopher Smith). New Haven: Yale University Press.
- Geach, P.T. 1966 'Plato's *Euthyphro*,' *Monist* 50, 369-82.
- Grote, G., 1865 *Plato and Other Companions of Socrates*, 3 vols. London.
- Irwin, T., 1973, *Plato's Moral Theory*. Oxford: Clarendon Press.
- 1995, *Plato's Ethics*. Oxford: Oxford University Press.

- Joseph, H.W.B. 1935 *Essays in Ancient and Modern Philosophy*. Oxford: Clarendon Press.
- Kahn, C. 1992 'Vlastos's Socrates,' *Phronesis* 37, 2321-58.
- 1995, *Plato and the Socratic Dialogue*. Cambridge: Cambridge University Press.
- Kant, I. *Groundwork of the Metaphysics of Morals* (tr. Mary Gregor 1998). Cambridge: Cambridge University Press.
- Kaplan, D. 1969 'Quantifying In,' *Synthese* 1968, 178-214.
- 1979 'On the Logic of Demonstratives,' in Peter A. French, Theodore E. Uehling, Jr., and Howard Wettstein (eds.) *Contemporary Perspectives in the Philosophy of Language* (pp. 401-12) Minneapolis: University of Minnesota Press.
- Kirwan, C. 1965 'Glaucón's Challenge,' *Phronesis* 10, 162-73.
- Kraut, Richard, 1984, *Socrates and the State*. Princeton: Princeton University Press.
- Mabbott, J.D., 1937, *Mind* 46, 468-74, revised and reprinted in Vlastos 1971b.
- Malcolm, N. 1959 *Dreaming*. New York: Humanities Press.
- Moore, G.E., 1903 *Principia Ethica*. Cambridge: Cambridge University Press.
- Nagel, T. 1968 *The Possibility of Altruism*. Princeton: Princeton University Press.
- Penner, T., 1971 'Thought and Desire in Plato,' in Vlastos (ed.), 1971a.
- 1973a 'The Unity of Virtue,' *Philosophical Review*, 82, 35-68.
- 1973b 'Socrates on Virtue and Motivation,' in E. N. Lee, A. P. D. Mourelatos, and Richard Rorty (eds.), *Exegesis and Argument: Essays for Gregory Vlastos, Phronesis*, Suppl. Vol. I: 133-51.
- 1987 *The Ascent from Nominalism: Some Existence Arguments in Plato's Middle Dialogues*. Dordrecht: Reidel (Kluwer).
- 1988 'Socrates on the Impossibility of Belief-Relative Sciences,' *Proceedings of the Boston Area Colloquium in Ancient Philosophy*, Vol. III (Lanham, MD): 263-325.
- 1990 'Plato and Davidson: Parts of the Soul and Weakness of Will', in David Copp (ed.), *Canadian Philosophers*, Suppl. Vol. 16, *Canadian Journal of Philosophy*: 35-

74.

- 1991 ‘Desire and Power in Socrates: the Argument of Gorgias 466a-468e that Orators and Tyrants have no Power in the City,’ *Apeiron* 24: 147-202.
- 1992 ‘Socrates and the Early Dialogues,’ in Richard Kraut (ed.), *The Cambridge Companion to Plato* (pp.121-169). Cambridge: Cambridge University Press.
- 1996, ‘Knowledge vs. True Belief in the Socratic Psychology of Action,’ *Apeiron* 29: 199-230.
- 1997a ‘Socrates on the strength of knowledge: *Protagoras* 351B-357E,’ *Archiv für Geschichte der Philosophie*, 79: 117-49.
- 1997b ‘Two Notes on the *Crito*: the impotence of the many, and “Persuade or obey,”’ *Classical Quarterly*, 47 (1): 133-46.
- 2000 ‘Socrates,’ in C.Rowe and M. Schofield (eds.) *The Cambridge History of Greek and Roman Political Thought* (pp.164-89). Cambridge: Cambridge University Press.
- 2002 ‘The Historical Socrates and Plato’s Early Dialogues: Some Philosophical Questions.’ In Julia Annas and Christopher J. Rowe (eds.), *New Perspectives on Plato, Modern and Ancient* (pp.189-213). Cambridge MA: Harvard University Press.
- Unpublished A ‘Towards a Prudential Reading of the *Crito*’ to appear in a volume edited by Vassilis Karasmanis.
- Unpublished B *Plato and the Philosophers of Language*.
- Unpublished C ‘Logic, Language, and the so-called ‘Socratic Elenchus’,’ de Vogel lecture at the International Plato Society meetings in Würzburg, July 2004.
- Unpublished D *Ethics and the Identity of Desire*. A revision of three Brian O’Neil lectures on the History of Philosophy delivered at the University of New Mexico, Albuquerque, March 6-8, 2000.
- Penner, T. and Rowe, C.J. 1994, ‘The desire for good: is the *Meno* consistent with the *Gorgias*?,’ *Phronesis* xxxix, 1-25.
- Forthcoming *Plato’s Lysis*, from Cambridge University Press.
- Perry, J. 1979 ‘The problem of the essential indexical,’ *Nous* 13, 3-21.
- Price, A.W. 1995 *Mental Conflict*. London: Routledge.

- Prichard, H.R., 1928 'Duty and Interest' in *Moral Obligation* (2nd edition), 1968. Oxford: Clarendon Press.
- Putnam, H. 'Dreaming and Depth Grammar,' in R.J. Butler (ed.) *Analytical Philosophy First Series*, reprinted in Putnam 1975b (pp.304-324).
- 1975a 'The Meaning of Meaning,' in K. Gunderson (ed.) *Language, Mind and Knowledge*, Minnesota Studies in the Philosophy of Science, Minneapolis: University of Minnesota Press, reprinted in Putnam 1975b (pp.215-71).
- 1975b *Language, Mind, and Meaning: Philosophical Papers*, Vol.2. Cambridge: Cambridge University Press.
- Quine, W.V. 'Notes on the Theory of Reference,' in *From a Logical Point of View* (pp.130-38). New York: Harper and Row.
- 1960 *Word and Object*. Cambridge MA: The MIT Press.
- Rawls, J. 1971 *A theory of Justice*. Cambridge MA: Harvard University Press.
- Reshotko, N., (Ed.) 2003 *Desire, Identity and Existence*. Kelowna BC: Academic Printing Press.
- Rudebusch, G. 2003, 'Socratic Perfectionism', in Reshotko 2003, 127-141.
- Ryle, Gilbert, 1931/32 'Systematically Misleading Expressions,' *Proceedings of the Aristotelian Society*, 32, 139-70.
- Sachs, D. 1963 'A Fallacy in Plato's *Republic*.' *Philosophical Review* 72: 141-158.
- Santas, G. 1979 *Socrates: Philosophy in Plato's Early Dialogues*. London: Routledge and Kegan Paul.
- 2001 *Goodness and Justice: Plato, Aristotle, and the Moderns*. Malden MA: Blackwell.
- 2003 'Penner, Plato, and Sachs on Justice and Happiness.' In Reshotko 2003, pp.95-107.
- Schofield, M. 2000 'Approaching the *Republic*' in C. Rowe and M. Schofield (eds.) *The Cambridge History of Greek and Roman Political Thought*. Cambridge: Cambridge University Press.

- Sidgwick, Henry, 1907 *Methods of Ethics* (7th edition). Cambridge: Cambridge University Press.
- Stampe, D.W. 1972 'On the Meaning of Nouns', D. Cohen (Ed.), *Limiting the Domain of Linguistics*. Milwaukee.
- Vlastos, G. (Ed.) 1971a *Plato*, Vol. II. New York: Anchor/Doubleday.
- 1971b, *Socrates*, New York: Anchor/Doubleday.
- Vlastos, G. 1991 *Socrates, Ironist and Moral Philosopher*. Ithaca NY: Cornell University Press.
- 1994 M. Burnyeat (Ed.) *Socratic Studies*. Cambridge: Cambridge University Press.
- White, N. 1979 *A Companion to Plato's Republic*. Indianapolis: Hackett.
- 1999 "Intrinsically Valued Parts of Happiness: Aristotle, Butler, and Mill."
Philosophiegeschichte und Logische Analyse = Logical Analysis and History of Philosophy 2:149-56.
- 2002 *Individual and Conflict in Greek Ethics*. London: Oxford University Press.
- Williams, B.A.O 1973 'The Analogy of City and the Soul in Plato's *Republic*.' In E. N. Lee, A. P. D. Mourelatos, and Richard Rorty (eds.), *Exegesis and Argument: Essays for Gregory Vlastos, Phronesis*, Suppl. Vol. I: 196-206.
- Wittgenstein, L., 1953 *Philosophical Investigations*. Oxford: Blackwell.

NOTES

¹ The notes to this paper explore many important further philosophical and exegetical details of the material in the main text. But the main text is designed to be read without significant reference to the notes. The notes most important to the main text are the present note, for its brief account of how I read the *Republic*, n. 30 on meaning and reference (as well as nn. 26, 28, 29, 31, 32, 37, 38, 54-58), n. 14 on desiring justice ‘for its own sake as a means to happiness,’ nn. 24, 25 on the historical Socrates, nn. 32, 53 on the so-called ‘Socratic elenchus’ with some remarks on Rawls’s approach to justice, as well as n. 43 on the peculiar nature of the egoism I attribute to this historical Socrates, by contrast with the notions of ethical egoism and ‘practical reason’ we find in Sidgwick. But, as mentioned, the main text is designed to stand on its own.

I refer in the main text to ‘apparent digressions.’ If we are to say what parts of the *Republic* constitute digressions, obviously we must ask ‘Digressions from what?’ And plainly this raises the following question: ‘Is the *Republic* primarily an ethical work? Or primarily a work of political philosophy? Or is it equally both?’ This question deserves more discussion than I can give it here. All I wish to claim here is that much of the *Republic* (at least Books II-IV, VI-VII, VIII-IX, and the second half of Book X) is under the tightest creative control in the service of the project of meeting the challenge of Thrasymachus, Glaucon, and Adeimantus just mentioned – that is, of answering the question, ‘Is the just *individual* happier than the completely unjust individual?’ and hence also answering the question this presupposes, namely, ‘What is justice (in the individual)?’ (Contrast the wholly political project that appears in the abstract of *Rep.* II-V that we find in *Tim.* 17b-19c – without a trace of this concern with individual justice and the happiness of the just). That is, I claim that the central project of the *Republic* is to focus on individual justice and individual happiness. On this view, the *raison d’être* of the construction and description of an ideal city is to provide a model by looking to which this project can be carried out. The focus on the individual as opposed to the political shows up in at least the following ways:

- (a) in Plato’s lack of interest in the justice of the individual *citizen of the ideal city*, by contrast with what does interest him: the justice of the city as a whole, and the justice of the individual *of no matter what city* (see the discussion of (JCITIZEN*) in sec. IV, as well as n. 20);
- (b) in Plato’s otherwise puzzling choice of ‘the city of pigs’ rather than ‘the good city’ as his ideal city (see below, n. 15);
- (c) in Plato’s apparent uncertainty as to whether it matters to the project of the *Republic* whether the ideal city ever could be realized (see below n. 51).

The reader should be aware that there are places (e.g., n. 51) where I shall argue that, in the light of what I am claiming is Plato’s primary interest in the ideal city *as a model for the just individual of no matter what city*, that, faced with a choice between saying

- (i) that the just individual (like the guardians of the ideal city) will look to the good of the city to which he or she belongs,

and

- (ii) that the Reason of the just individual (like the guardians of the ideal city) will look to the good of the whole of which it is just one part (so that the just person looks to the good of his or her entire soul).

Plato will opt for (ii). Had Plato’s primary interest been in utopian political philosophy, concern with the *citizen* of the just state would have led us to opt for (i). From the present point of view, this involves the tail wagging the dog. The correct choice here is (ii). These are very large issues, and I do not pretend to have settled them

here. I am just making clear how large the stakes are, and where I stand on the issues.

This said, I would certainly not deny that Plato himself – unlike Socrates – had enormous interest in details of political arrangements. This is especially obvious in *Letter VII* (if that is to be trusted – and its picture of Plato the young man is surely quite plausible), in the *entirely politically-focused* précis of *Rep.* II-V in the *Timaeus*, and in the *Laws*. But it also leaks out willy-nilly all over the place in the *Republic*.

² Sachs usefully cites passages where (A) Thrasymachus, and (B) Glaucon and Adeimantus seem to explain justice in terms of the non-performance of acts of certain kinds. Thus (A) at 343d2-e7, Thrasymachus represents the just person as the one who always ‘gets less’ in a temporary business partnership, who pays more when it is a question of tax assessments, who ‘gets less’ when there are distributions, and whose personal affairs deteriorate when holding public office, at the same time as he or she also alienates his friends and relatives through not helping *them* ‘get more’ from this office. Similarly, in Glaucon’s account (360b5-c3), the just person is precisely the one who does *not* do what Gyges with his magic ring, *would* do. Gyges would be unable to keep his hands off the property of others; he would enter into just any house and sleep with whomever he wished; and he would kill or release from prison whomever he wished. In Adeimantus’ account (366c6-d1), the just person is either the one who scorns the doing of injustice – this as a result of a divine nature in him (we might call this ‘the just man as miracle,’ a quite ordinary notion of justice, in Greek antiquity as in the modern era: compare also Thrasymachus’ ‘simpleton’ at 348c11, 361b5-8 and earlier at 343d3) – or he is the one who, in a sop thrown to Socrates, keeps away from such deeds as a result of having achieved some form of knowledge. In short, the just person is the person of such a sort that, whatever his inner state may be, he does not do such unjust deeds as stealing, forcing himself on others, killing, illegally releasing prisoners, and the like.

³ Sachs himself does not describe his non-performance conception of justice as a *moral* notion. But it seems plain enough that Sachs supposes that the ‘justice’ he introduces in terms of the non-performance of acts of certain kinds actually *means* the ordinary moral notion. Thus he speaks of ‘a list of acts ... incompatible with justice and as ‘constituting’ injustice’ (5th para.); of a conception that ‘precludes’ behavior commonly judged immoral or criminal; of ‘ordinary or vulgar canons of justice’ (main text, 2 paras. up from n. 21); of showing that Platonic justice ‘entails’ ordinary morality (five uses of ‘entails’ in the last 7 paras. of Sachs’s article); and of whether certain theses ‘exclude the possibility of’ the Platonically just committing what would ordinarily be judged immoral acts (6th last para.). It seems a safe assumption here that all this talk of ‘incompatibilities,’ ‘constituting,’ ‘precluding,’ ‘excluding possibilities,’ ‘entailments’ and the like adverts not to talk of truth and consequence merely by virtue of logical form, but rather to talk of truth and consequence by virtue of logical form *plus rules of meaning*, as determined by ‘criteria.’ (See also the talk of ‘criteria’ in the fourth paragraph of Sachs’s paper). Hence my suggestion that Sachs thinks ‘justice’ *means* the ordinary moral notion.

Of course it is one thing to *list* acts, and even kinds of acts, another to provide a function or conception which would *generate* such a list. Perhaps this is why Sachs concedes (his 5th para.) that the nonperformance of certain kinds of acts does not ‘exhaust the ordinary conception.’ We seem driven, once more, to the view that what Sachs has in mind as the ‘ordinary conception’ is a certain sort of moral-rules conception of justice – filled out theoretically in terms of the sort of intuitionist, rule-based theory of morality given classic expression by Sidgwick 1907 [1874]. This would bring Sachs’s complaint against Plato rather close to that in Prichard 1928, which also accuses Plato of a ‘change of subject’ not altogether unlike that in Sachs’s accusation. (I had originally proposed to include a study of Prichard’s charge in the present paper, but that proved unwieldy. I shall return to such a study elsewhere).

⁴ Sachs draws here on 442e1. His misinterpretation of this passage is discussed below, n. 27. This misinterpretation does not affect the claim that Sachs *thinks* the conception of justice shared by Thrasymachus, Glaucon, and Adeimantus is the ordinary or vulgar (moral) notion.

⁵ Note how Sachs in this passage glosses *Plato’s* reference to justice as ‘justice as he conceives it to be,’ while

not so glossing Thrasymachus' reference to justice. This shows the effect of the presupposition in Sachs that the real nature of justice (or at any rate, all there is to justice) is just what the 'vulgar conception' of justice conceives it to be. Sachs sees no need, therefore, also to gloss Thrasymachus' reference to justice as justice *as conceived by Thrasymachus*. I should add, however, that this does not undercut the view for which I argue in sec. V below, that whatever Sachs may *presuppose*, all he thinks it necessary to *use* in his argument against the Socrates of *Rep.* II-X is that Socrates fails to address himself to whatever conception it is that Thrasymachus employs in his challenge.

⁶ See below, n. 30 for brief statements on how I understand the notions of meaning and reference, as well as on how I understand the thesis that meaning determines reference (as well as various liberalized versions of the thesis that meaning determines reference, e.g., the thesis that meaning plus 'context' determines reference). The usual view, under which I claim Sachs's view falls, is that if there is reference, then there must be ways in which the reference is determined which are known in advance – say, by our 'knowing the language.' By contrast, the view I defend briefly in n. 30 has it that, since reference *is* often successful, reference *is* determined, and if so there will be *ways* in which it is determined; but (as causal theorists, for example, have maintained) the reference need not be determined in any way of which we have knowledge in advance. The remarks about 'blank check' theories two paragraphs down give a preview of my opposition to the idea of advance knowledge of ways reference is to be determined. See also my remarks on the 'quantifier shift' at the very end of the paper. For other examples of commentators who, in one way or other, follow Sachs, see n. 37 below.

⁷ Of course, for believers in meanings, expressions quite different in meaning *can* have the same reference, e.g., 'The second planet from the Sun' and 'The Morning Star.' My point here is that on the assumptions with which Sachs operates, he will reasonably deny that the same reference could be determined by (a) 'justice' in the meaning Thrasymachus employs (non-performance of acts of certain kinds) and (b) 'justice' in the meaning employed at least sometimes in the *Republic* (as a form of psychological well-adjustment according to Sachs). For in Sachs's view (as indeed in the view of Joseph 1935), it is possible for some well-adjusted people not to refrain from performance of acts of the 'immoral' kind and for some people who refrain from 'immoral' acts to be psychologically ill-adjusted. So Sachs will maintain that the difference in meanings or conception here assures difference of reference.

⁸ Sachs does not use the word 'meaning.' But it is pretty clear that this is what he has in mind: see n.3 above. On concepts and linguistic intuitions, see again nn. 37 and 38 below.

⁹ It is true that Thrasymachus still holds that the real nature of justice he is thinking of is something that corresponds to his conception. But that does not tell either way on the question which of (i) and (ii) he would opt for if – as will inevitably happen – he is forced to choose between them. For more on this point, see n. 28 below. Notice that (i) is not about the advantage of the stronger *tout court*.

¹⁰ I would agree that most Greeks thought that justice was a *moral* notion, not exhaustible by the psychological accounts offered by Plato – or indeed by the historical Socrates (the Socrates, roughly, of the early dialogues: see next note). See, for example, sec. VI of my 2001. Hence, I find much to agree with in White 2002 on thinkers *other than* Socrates and Plato. Unlike White, however, I see Socrates and Plato as presenting a radically new and non-moral approach to ethics. Socrates' accusers had no very discriminating way to get a handle on Socrates. But they were right to think that what Socrates was doing was a radical departure from anything that had occurred up to that point. See my neo-Prichardian scenario in the forthcoming Blackwell's *Companion to Ancient Philosophy*. (On Socratic justice as goodness at getting one's own happiness, see nn. 16, 23-25, 42-44 below).

¹¹ For a beginning on how I view the early dialogues and the question what in early Plato is historically Socratic, what is not – on which I shall have a good deal more to say in upcoming publications – see my 2001. There I make clear why I think we can only say 'roughly' that the early dialogues are Socratic (so that, for

example, the *Gorgias*, especially the Callicles part and the myth, and Book I of the *Republic* (n. 25 below) both contain some non-Socratic elements, and the *Symposium* contains both distinctively non-Socratic arguments, concerning the Forms and flux (as does the *Cratylus*), as well as distinctively Socratic elements, concerning desire for the good. In the sequel, the ‘roughly’ should be understood throughout, even though I omit it. Of course the *Phaedo* is also stylometrically early; and much of the philosophical part of that dialogue sounds distinctly non-Socratic.

¹² Again see n. 23 below, with the fuller discussion in nn. 42 and 44 of the ‘purely factual’ in Socratic ethics.

¹³ I myself, it will be apparent by now, have considerable doubts about what justice is after all – at any rate, if justice is also to be taken to be a virtue. Half of the time, I tend to think that justice is what most people – or at any rate most philosophical opinion-leaders – think it is, e.g., what we find exemplified by most procedurally correct trials and administrative decisions; and I tend to think that the qualities such a notion produces in people do not amount to virtue. [See, for example, the ‘justice’ referred to at *Gorg.* 464b8, c2, 465c5 – the justice of the law courts – if Dodds is correct to read *dikaiosunē* here, with mss. other than F, rather than *dikastikē*, which is the reading in F. Dodds notes as support for his view that *dikaiosunē* is the unique reading in all mss at 520b3. The play with justice and rhetoric at 454b7, 459c8-461b2 (in the first part of the dialogue where Gorgias is the principal interlocutor) tends to support Dodds’s reading]. I should admit, however, that my experience with procedural justice is rather limited, being restricted more or less to departmental faculty meetings. The other half of the time, I tend to think – with Socrates and Plato (as I see their views) – that justice *is* a virtue but not at all what people think it is. (See below, n. 50). Now, it surely cannot be wrong for an interpreter – myself or anyone else – to look at the *Republic* from the point of view of such doubts. (Especially if, as in the present case, the interpreter fancies he came by those doubts by reflecting on the views of the Socrates of roughly Plato’s early dialogues: see esp. secs. VIII-IX below). At any rate, this is how I approach the *Republic* in the sequel.

¹⁴ It seems desirable here to make clear that the present passage brings no support whatever for the idea that the reference to justice ‘in itself’ is a reference to justice aside from *all consequences whatever* – that is, a *moral* conception of justice as opposed to one based upon such advantages from justice as accrue to one prior to the social contract and the rewards of reputation (where such rewards represent yet further advantages). It is true that Plato speaks of justice being desired *in itself* – *di’ hautou* at 358a1-3 – and also of its being embraced for its own sake (*auto ... hautou heneka* at 357c1-2, b6), as opposed to its being desired for the sake of its consequences. And it is also true that the reading of the suggestion that justice is ‘desired for its own sake’ that comes naturally to modern moral philosophers (compare Mabbott 1937, Kirwan 1965) would give it an absolutist moral reading: roughly justice is desired on purely moral grounds, and without regard to *any* consequences for our happiness. But this does not appear to be what is involved in the present passage. For as Irwin 1995, 190, decisively points out, both things desired ‘for their own sake’ and things desired ‘for the sake of consequences’ fall under ‘desired by the man *who is to be happy*.’ It is my view that, as Prichard 1928 remarks, at 1968, 205, anything desired by the man who is to be happy (358a2-3, compare *Gorg.* 507e1, *Euthyd.* 280d5-6), i.e., as a means to happiness, can only involve a hypothetical imperative, and hence cannot involve morality (where the imperative would have to be categorical). Irwin, on the other hand, hopes by other argumentation (to be explored later in this note) to restore the *status quo ante*, wherein ‘desiring for its own sake’ *is* moral, and so does not involve the hypothetical imperative mentioned.

The result of these considerations is that what is needed here as an explanation of ‘desired for its own sake’ is something like *desired for its own sake as a means to happiness*. (The presence of this idea of being desired for its own sake is clear enough from a passage in the *Euthydemus*. Thus, at 281d2-e5 wisdom is said to be, of all the good things there are, the only thing good in itself. But when we look at the preceding context, esp. 278e3-279a3, 279a5-c2, we see that all these supposed good things, including wisdom are, in effect, characterized as *good as means to gaining happiness*. Hence wisdom is – odd as it sounds to modern Kantian ears – the only thing good in itself as a means to happiness).

But how are we to understand such an idea? Let us begin with *things desired not for their own sake but for their consequences as means to happiness* – the third and clearest of the following trichotomy:

for own sake only / for own sake and for consequences / for consequences only.

It is pretty plain that the content of ‘for consequences only’ (i.e., ‘not for its own sake’) is that things of this sort are in their initial aspect irksome (357c5-d1, with 358a4-6, b5-6, 364a1-4) – exercise, by itself, is a bore, being honest, by itself, is a bore – but they *indirectly* bring benefit (say, via reputation in the case of honesty, via endurance and success in battle in the case of exercise). The idea must therefore be that to desire justice *for its own sake* (= *for its own sake as a means to happiness*) is to desire it for the benefits it brings in its initial aspect, i.e., *directly*. This idea receives further support in the idea, prominent in this passage, of desiring the inner power (*dunamis*) justice for what it accomplishes *by itself* (358b5, 366e4-5, 367b2-5; and c7, d3 when taken together with d2, e3-4) vs. what it accomplishes via reputation and so forth (357c1-d2, 358a4-6, 363a1-5, 367b2-c1, d2-e5, 392c2-4, 612a8-d9, 613b9-c6, and compare 361a5-d3). [The talk of the rewards of justice gotten by means of one’s having the reputation for it suggests that justice, unlike gymnastic training or medical cures, never brings any rewards *other* than those rewards that result from a reputation for justice. This suggestion has caused interpreters trouble, since it is not explicit in the formal statement of the trichotomy at 357a4-d2. But we should recall that although Socrates *accepts* the trichotomy, the point of the trichotomy from Glaucon’s point of view is to represent the irksomeness of justice from the point of view of Thrasymachus, who will have supposed that the only thing justice is good as a means to *is* indeed reputation. So we should not be surprised that this is the only example of things good because of their consequences which is employed in the discussion of the Thrasymachean social contract]. Something like this conception of desiring justice ‘for its own sake’ is to be found in Foster 1936, Sachs 1963, White 1979 (though without any explicit employment of the present idea of being good for its own sake as a means to happiness).

I find unconvincing the attempt of Irwin 1995, 190-191 with 379, n. 20, to construe the idea of the happiness which is desired ‘for its own sake’ and which the *dunamis* of justice brings about in us in *conceptual* rather than *causal* terms – *dunamis* as formal cause rather than as efficient cause of the happiness produced in the agent. According to this construal, Plato is saying not ‘justice causes us to be happy’ but ‘being just is *part of the concept* of happiness.’ (Irwin’s reading restores the morally absolutist reading of ‘in itself’ by, in effect, building the *moral* meaning of ‘justice’ into ‘happiness.’ Thus it also makes the ‘eudaemonism’ Irwin attributes to Plato a *moral* eudaemonism – an exceedingly resourceful move. My resistance to such a move is both philosophical (I do not find such appeals to the ‘conceptual’ *philosophically* persuasive: some of my reasons emerge below), and exegetical (the entire effort of the present paper being precisely to get away from just such use of meanings, conceptual connections, and so forth in interpreting the *Republic*). None of this is to deny that Irwin does as good a job as can be done by a meaning theorist in defending the use of meanings in the *Republic*. (I have profited here from many conversations with James P. Butler).

¹⁵ Not a state of ideal citizens, since it is instead the ideal ‘fevered’ city, containing people as they *are* (with the desire for luxury somewhat indulged: 372d4-374a2). Why does Plato not have Socrates examine the ‘true’ city (372e6), with citizens who have not been corrupted by the taste for luxury (371e9-372d3, and compare 539d1-544a1)? Indeed, why first introduce the city with luxury (even introducing poets at 373b), and then purge it of at any rate some forms of luxury, especially certain kinds of poets and poetry, music, and so forth (starting at 378b and climaxing at 399d), before saying what justice is in the (first corrupted, then) purged city? So I ask again: why not stick with the true city?

I suggest that the reason has nothing to do with what might be ideal in other respects, and everything to do with the psychology Plato thinks holds of all people. The point is that to deal with justice in the soul, you have to deal with an appetitive part that is by nature unruly and, without opposition from the rational part, likely to bring about actions of the sort Aristotle would later call irrational voluntary actions. (Here the big difference shows itself between Socratic intellectualism and the decidedly limited rationalism Plato displays in the *Republic*: see sec. VIII below, esp. para. 3 ff.). Justice must show itself in the degree of control Reason and the

spirited part have acquired over such an element, and in the degree of agreement they have won from such an element. The craving in some parts of the population for luxury (373a1-b1) is merely the civic analogue to unruly appetite in the soul. It is true that none of this suggestion shows up at the point in the text where Socrates turns from the ‘true’ city to the fevered city. What shows up, in some detail, where Socrates turns from the ‘true’ city to the ‘fevered’ city, is the origin of war (373d), and a concern with the evolution of the military (373e9ff). But concern with control (in conformity with Reason) may surely show a concern with the unruly lower element (the origin of war in the irrational parts of the soul: see the references to the military keeping internal order in (R2) below). Such reflections tend also to reinforce a suggestion which will become central to the present paper, namely, that Plato regards as essential to his getting justice right that he get his psychology right, as well as the suggestion already made in n.1 that most of the *Republic* is under the tightest creative control of a vision concerned primarily with the justice in the individual psyche.

¹⁶ See Santas 2001 for excellent argument that the notion of function is as much in operation in Book IV as it is in Book I. I take it that, for Plato, the function of each of the three classes flows from two things: (a) the abilities of people in those classes, and (b) the happiness in the city that the political theorist attempts to devise, analogous to the happiness of the individual. Thus for me the function of a class in the city looks to the good of the city, so that the function of a part of the soul of an individual should look to the good of that individual. (This is explicit in the case of the rational part – see [R3] immediately below.) The effect, for *individual* justice, of concern for function in the *Republic* is thus, by my lights, that function is directed at *the agent’s own good or happiness*.

The idea here may be illuminated by some more Socratic considerations that do not turn on parts of the soul. At *Rep.* I.352d8-354a9, the function of the *soul* is to live, and the *virtue* of the soul is to live well, and so be happy. So the good person is the person good *at* living, i.e., at being happy. (Compare *Gorg.* 507b8-c5, and on a good soul being a soul good *at* some-thing, see *Hipp. Min.* 373c ff, as well as my 1973b). The result is that the function of *my* soul is to make *me* happy.

This idea about function is too egoistic for many interpreters of Plato, e.g., Santas 2001, ch. 3, who want the good involved to be not *my* good, but good in general. I would agree that the knowledge of the good and the bad which virtue consists in is *general* knowledge. (Just so the cobbler’s knowledge of shoes is *general* knowledge of shoes). But how this general knowledge is applied by the good person is to his or her own good. (Just so, the cobbler applies his or her general knowledge of shoes to make the shoes *in his or her shop*. He doesn’t try to make *general* shoes). Hence my view that Socrates’ intention here is to say that the function of the good soul is to make the person in question happy.

I see the connection of virtue and function with knowledge in Socrates as follows: The knowledge which is virtue is the knowledge or science of (human) goods and bads – i.e., a knowledge or science of happiness *in general* and not just of *my* happiness. (A good thing too, if my happiness is wrapped up in the happiness of others around me). But by virtue of the fact that all my actions proceed from the desire for my own good (see on Socratic Intellectualism in sec. VIII below), we get that the way this general knowledge or science gets applied to me is in aiming at my own happiness. For, as we’ll see below, the situation with this science of happiness in general is the same in Plato as it is in Socrates, except that Plato thinks that to get to act on the relevant knowledge, one has to overcome obstacles from the appetitive part – obstacles both to the acquiring of the knowledge and also to acting on one’s knowledge. See further below, the 2nd para. of n.51 below on the Form of the Good as general, and the claims of a number of scholars that the fact that the Form of the Good is general shows that Plato is no egoist.

¹⁷ 427c6-d1 with d4-6 goes from considerations of the city and its justice to the question whether that justice *in the individual* makes the individual happy (independently of whether or not gods or men notice that justice). See also 434d2-435a3, 435a5-c5, 441d1-442a2, 442d2-3, 443b4-5, 444a4-6, as well as the passages from Books VIII-IX cited in n.22 below. On there being a single identical structure to both ideal city and just soul, see the discussion, towards the end of sec. IV, of the mistranslation of the title of this dialogue, *Peri Politeias* as

‘The *Republic*.’

¹⁸ The passage in which Plato makes this argument (445a5a5-b4) runs as follows:

It appears to *me*, Socrates, that the inquiry [into whether doing just deeds and being just profits us, whether or not one is noticed in this] becomes ridiculous from this point on (*êdê*): If life does not seem worth living with the nature of the *body* corrupted (even if one has all manner of food, drink, wealth, and political office), will it still then (*ara*) be worth living with the nature of that very thing we live *with* [viz, the soul, here, as at *Crito* 47d4-5, unnamed] disturbed and corrupted (even if one does whatever one wants except the very thing which will release one from vice and injustice and acquires justice and virtue) – now that (*epeidêper*) it has become transparent to us what each of these [justice and injustice] is?

A life ‘worth living’ (*biôton*) here, as at *Crito* 47c-e, I take to be a sufficiently happy life. For precisely the point of the present passage is to show that the unjust life is unhappier. I am thus supposing that the position of Mabbott 1937 and Kirwan 1965, already cited in n.14 above, shared also by Irwin 1995, 252-3 – that health is desired for its own sake *as opposed to* its facilitating a happy life – is implausible and cannot be made out. (There is no occurrence of any such phrase as ‘for its own sake’ in the passage – let alone an occurrence of ‘for its own sake’ read *morally*. Notice also 444e7-445a3 on ‘whether it profits us to act justly ... whether or not one is noticed doing so.’ ‘Profit’ would have to be ambiguous on this view, giving us *conceptually* profiting for justice, *causally* profiting from being noticed. This seems rather forced).

¹⁹ I am indebted to Sara Gavrell and, especially, Andrea Veltman for convincing me that sec. I of Cooper 1977 has this exactly right. Prior to this, I am afraid, I had supposed that justice in the *Republic* was meant to be something that the ordinary citizen of the just state would have, without having knowledge. That of course requires the shuffle of supposing that justice in Book IV does not require that knowledge be in one’s own rational part, but merely that knowledge be in the rational part of one’s commanders. (More on this in the next paragraph of the main text). Having said this, I must also record that there is much in the rest of Cooper 1977 with which I do *not* agree. See again the 2nd para. of n. 51 below.

²⁰ It will be useful to follow Cooper by approaching the topic of demotic virtue (*dêmotikê aretê*) in the *Republic* via citizen courage (*politikê andreia*) at *Republic* 430c1-5. In the passage leading up to talk of citizen courage, Socrates is giving his account of courage in a city (429a-430c). The thought here seems to be this: A city is courageous if its military part preserves the belief that what is to be feared are the various things that, in the military class’s education, the law said were to be feared. The preserving of these true beliefs is described in terms of the beliefs being fast-dyed into the psyche, so as not to be washed out by pleasures, pains, fears, and desires. Socrates first says that this is his account of courage in the city, as if he is saying that

a *politikê andreia* is *genuine* courage (as it shows up in the city).

But then he retrenches. He retrenches when confronted with slaves and animals who happen to have the right belief: since those right beliefs are not from education, slaves and animals are not courageous at all (430b6-9). Why then does Socrates now proceed to back off saying that the military have genuine courage and further retrench to saying, in effect, ‘Well, to be strictly correct, the courage had by the military is *citizen* courage’? (*Apodechomai Kai gar apodechou ... politikên ge, kai orthôs apodexei.*) The point here appears to be, by contrast with (a), that

b *politikê andreia* is not *genuine* courage, or even a species thereof, but an inferior form of courage.

Why the retrenchment from genuine courage to mere *citizen* courage? Plato is talking about the courage of the military citizens of the ideal state: so why isn’t the courage of a citizen of the military class genuine courage,

identical with genuine courage in an individual? (See esp. 429a8-b4, the point of which cannot be, however, that the *guardians* – members of the intellectual class – are *not* courageous).

All Socrates tells us in his explanation of the qualification that the military in the ideal state have *citizen* courage (430c4-5) is that he can give a fuller account of courage elsewhere – a remark that can only be referring forward to 504aff, where it is made clear that nothing said in Book IV to be a virtue can *be* a virtue without knowledge of the Good. But now the courage of a member of the military is not based on that member's knowledge of the good, but on an education which is really more a molding of one's character (500d4-8) *without* the imparting of knowledge, than a matter of knowledge. It is more a product of habit than philosophy or thinking (619c6-d1, 518d9-519a1; and notice that true opinion without knowledge is blind: 506c6-10). Hence the courage of a member of the military class is not genuine virtue, but citizen virtue. The use of the word 'citizen' at 430c3, and 'demotic' at 500d8 together recall the reference at *Phaid.* 82b1-3 to

demotic and citizen virtue which they call temperance and justice when they [temperance and justice] come to be from habit and practice rather than from philosophy and mind. (See also 68d-69c).

Thus Cooper's argument seems to me conclusive: since courage in a city is the city's having its middle part preservative of true beliefs given to it by *its own* rational part, it follows that a citizen member of the military in a courageous city will preserve true beliefs without that citizen member of the military having knowledge in *his or her own* rational part. So these soldier citizens won't strictly have courage – only what we can call citizen courage. As for the spirited part of the military acting on knowledge supplied by the guardians (*Rep.* 590c2-591e3), I agree with Irwin 1977, 329-30, that such action is not virtuous.

In the *Laws* we have an interesting difference from the *Republic*. For the *Laws* also has a distinction between demotic virtue (this time *dēmôdês aretê*) and virtue proper (710ab). But now it is beasts and children that are allowed to have the demotic virtue, and those with knowledge *or true belief* are the ones allowed to be properly courageous. Hence in the *Laws* (720b-722c, with 857d-858a) we have *this* contrast: demotic virtue involves the slavish obeying of orders, while it is sufficient for genuine virtue that one merely undergo true *persuasion* (even without knowledge) as to virtue. Possessors of demotic virtue, like slave doctors, blindly follow the rules. The genuinely virtuous, like free doctors, have some intercourse with true belief (in the form of listening to the *preludes* to the laws). Mere true belief no longer downgrades one to demotic virtue. Thus Plato has a different theory of virtue in the *Laws* – in its idea that bureaucrats (those who don't have the full picture, and who in any case don't have full discretion) can nevertheless actually themselves be virtuous. My own life-experience with bureaucracies does not encourage me to view this picture in the *Laws* as anything but wildly optimistic. [The same will hold for bureaucracies based upon the idea of true belief without knowledge had by the military class in the *Republic*: see my 2000, n.16. With the notion of bureaucracies and intellectual material handed down to them from those with knowledge, we get what we may call 'bureaucratic' (i.e., not fundamentally comprehending) belief and 'propositions' to be the entities which the leaders know and the lower bureaucrats truly believe].

²¹ Williams 1973, appealing to 368d, 435b, 434e, argues that this structural identity is affirmed on the basis of a semantical theory according to which any word whatever must always stand for the same thing or the same structure no matter what the context and no matter what the word modifies. The passages Williams cites certainly show the structural identity in question. But is Plato declaring this identity of structure on the basis of any simple-minded semantical theory of the *unum nomen, unum nominatum* sort? My remarks below about the care with which Plato gerrymanders the structure of the city that is to be the just city (a city that does not occur in nature) should suggest to us that this is not so. Thus I do not believe Plato would have made the point at *Rep.* 434c7-435c6, or the analogy of small letters and large letters at 368d4-369b4, if he were not antecedently convinced that *in fact* human goodness depends crucially on the structure of one's psyche, and that, in fact, precisely that structure can be modeled in the just city we go on to construct: compare the similar point I make in n.15 above. (Plato was given to seeing these large structures. For example, consider the art of ruling in the *Statesman*, esp. 258e-259c, where Plato identifies the science of rule over slaves with the science of rule over a

household, and with the science of rule over a city. This identity is *not*, I suggest, a consequence of some argument of the *unum nomen unum nominatum* type). What I am saying, in other words, is that Plato is not arguing *from* the semantical theory, but merely stating that we have in the structure of individual *justice* a case where such a [preposterous] semantical theory would coincidentally give the same result. I believe we should also treat merely contextually, and in a more restricted way, the semantical theory *unum nomen unum nominatum* that we find at 596a5-b1. For it applies well enough to tables and beds. (Analytical philosophers – even the very great amongst us, which Williams certainly is – do us no favors in recklessly and unnecessarily diagnosing Plato’s argument by attributing to him such mindless semantical ‘models.’ I make similar remarks about Williams’s diagnosis in terms of the part/whole rule in the next note).

²² When I say that the account of justice in the city is designed to image justice in the individual psyche with its rational and irrational parts of soul, I do not wish to deny that the political material involved in the *Republic* has some influence on what Plato says about the psyche. The beginning of the *Timaeus* (see the preceding paragraph in the main text, as well as n.1 above) and the accounts of communism and the equality of women in Book V of the *Republic* show Plato had these interests in political arrangements. Indeed, in Penner 1971, I went so far as to argue that given the presence of a very substantial argument for the existence of two parts of the soul, and no argument of even remotely comparable merit for a third part distinct from Reason and Appetite, we are driven to the assumption that Plato’s impetus to a spirited part flows from his decision that he needed a military class for his ideal state – bureaucratic intermediaries between a rational class and an appetitive class – able to take orders, and to accept faithfully beliefs passed down by reason, and with the strength to enforce those orders on the appetitive class. This seems to suggest that the psychological is in the service of the political, contrary to my suggestions in n.1 above. What I now think is that the choice to have a military class in the model of the city itself has something of a psychological explanation. For if we look at the implausible arguments on the curative social effects of punishment in those who have acted unjustly that we find for the first time in the *Gorgias*, not only in Socrates’ ironic treatment of Polus, but also in the Callicles section (470a ff, 493a ff, 505b ff), we seem to find claims less directed towards a social theory of punishment (*kolazein*: 476a-477a, 478a, 479a, 527c) as a way to improve the souls of citizens than towards what will be for many a rather more persuasive theory about the benefits to the individual soul of disciplining (again *kolazein*) one’s bodily appetites in order to avoid the vice of *akolasia* (self-indulgence, or more appropriately in a translation – as a number of translators and commentators have alertly seen – ‘indiscipline,’ cognate with *kolazein*: see 491e, 505b-c, 507c-e). In this case – and what we have here is a foreshadowing of the city-soul analogy of the *Republic* – it looks as if the primary focus of interest is the individual psyche disciplining its bodily appetites. That is, it looks as if the (admittedly much more extensively discussed) interest in that disciplining of those who do injustice (which we know under the name ‘punishment’) as a tool for improvement in virtue is intended more as a model *confirming* the material on discipline in the souls than as a self-standing contribution to social policy discussions. This suggests the possibility that it is not so much that

- a the belief in *kolasis* as a tool for making citizens better society has generated a belief in *kolasis* as the disciplining of unruly appetites

as it is that

- b looking around for a model for talk of disciplining appetites, he has hit upon a model in the form of the more familiar practice of punishment that he thinks will help him get across such ideas as that of disciplining one’s appetites, and indeed of being master of oneself (491d-492a, *Rep* I.351e-352b).

Let us return now to the *Republic*, where Plato has also made explicit the further – and quite different – point that unruly irrational appetites can *directly cause action independently of Reason*. (One *could* suppose that there was need to discipline appetites merely to stop them from seeming so pressing that Reason would conclude that it would be good to satisfy them. On such a picture it would not have to be the case that these appetites could bypass calculations of good altogether, and directly cause action, in the way they do explicitly in Book IV of

the *Republic*. Once appetite can directly cause actions, the intellect will need more than knowledge to do the right thing. That is, Reason will need to have available to it the ability to coerce irrational appetites).

Now, this claim by itself – that appetites can directly cause action independently of Reason – hardly justifies the claim that we can infer a third part of the soul besides Reason and Appetite – a third part that does the coercing of appetites. For why shouldn't Reason, which, after all, does have desires of its own (*Rep.* IX.580d-581a), itself have the ability to coerce? But if in the *model* of the ideal city which the *Republic* constructs we are to have some coercing (and punishing) of the more unruly citizens, will it not be natural, *within such a model*, to suppose that the rulers will need police to discipline the unruly? So in this case, the model may have so far seduced Plato from determining the political in terms of the needs of the psychological, that the political made him miscalculate in his psychology. Hence I no longer wish to leave the impression my earlier piece of 1971 may have left, that I would say that Plato gerrymandered his psychology to fit the needs of a supposed independent interest in utopian political philosophy. I now wish to make clear that all the *Republic* itself justifies is that, on the question where the coercion and disciplining of irrational desires is to come from, Plato may have been unduly influenced by his model.

My suggestion here is continuous with the remarks in n.15 above, that we construct a model which is a fevered city, rather than a good city, so that we will have in the model an element that will coerce, discipline, and persuade an unruly irrational part. This invites the suggestion of a third element in the city – and so also in the psyche that the city models. Purely at the psychological level, nonetheless, Plato might well have realized, as did Aristotle, that he, Plato, needed only two parts, the rational and the irrational (*Rep.* X. 602c-608b and even IV.430c8-432a9, *Phaidr.* 237b-238c, *Leg.* I, *EN* I.13, *MM* I.i.8).

I cannot agree with Williams 1973, para. 3 ff that 435e1-436a3 shows Plato endorsing the view that a city is *F* iff all or most of its citizens are *F*. (Williams's whole/part rule). The passage says, true enough, that there are the same elements and characteristics (*eidê kai êthê*) in individuals as in cities: if a Thracian or Scythian city is spirited that will be because its individual citizens are spirited, and similarly for Athens being an intellectual (*philomathes*) city, and Phoenicians and Egyptians being money-loving. But we need to attend here to what Plato's point is in making the remark. The point of the remark is to set up the main question of the passage (436a8-b3), whether when we speak of a person being a lover of learning or spirited or appetitive, we are speaking of such a person being so because *all* of the self is that way, or because *one part* of the self is that way. (The civic analogue would be whether for a city to be money-loving or spirited or intellectual was for everyone in the city to be that way, or only some people in the city – in the case of being intellectual, we may add, even a small minority. This of course right away contradicts the 'all or most' of Williams's whole/part rule). The point Socrates is making here may then be of the following form:

(A) You might think that cities are spirited because all or most of their citizens are spirited;

on the contrary,

(B) cities are spirited because their military class is spirited.

And we find eventually that it is (B) and not (A) that is in question. In the circumstances, to extract a serious Platonic principle from (A) hardly seems justified. ((B) is of course Williams's 'predominant section' rule).

Even at that, (B) does not generalize from being spirited to being courageous. For we have just seen that cities *aren't* courageous by their military part being courageous, since the military part does not have genuine courage, only citizen courage. This is the point made above that there is no argument from (JCITY) to (JCITIZEN). If, with Williams, we want a generalization of the point being made in 435e1-436b6, it will be at best that a city is *F* (e.g., courageous, just) because of what all its citizens *do* (e.g., all do their own) or because of what some class of its citizens *does* (e.g., preserve beliefs imparted by other citizens) not because all the citizens *are F*, or because some class of the citizens *is F*.

It is true that, switching to Books VIII-IX, we also find at 544d6-e2 (which Williams also cites as an example of the part/whole rule) the argument that there are as many kinds of men as constitutions (*politeiai*), since the constitutions come from the characteristics (*êthê*) of the citizens, so that if there are five kinds of city, there will be five constitutions (*kataskhukai*) of individuals. But in *this* context, Plato's concern is not, as it was in Book IV, deriving characteristics of states from the characteristics of its citizens, but with deriving characteristics of *individuals* with the kinds of souls whose constitutions are modeled by cities with the four degenerate constitutions. The point is not to examine the *citizens* of the four degenerate constitutions, but those individuals of no matter what city whose souls have the constitutions modeled in the four degenerate cities. Plato tells us at 544a1-9 that the express aim of Books VIII-IX is to examine the faults of the four degenerate constitutions (*politeiai*), timocracy, oligarchy, democracy, and tyranny (544c1-d3), in order to show which *man* is happiest of the four corresponding men compared with the just man (544a5-9, notice also 544e7-545b1, once more concerning city and corresponding psyche, *not* city and corresponding citizens of that city). Again, 545b3-c7 correctly describes the structure of the rest of books VIII-IX in terms of speaking first of city, then of corresponding individual psyche, and not in terms of city and corresponding *citizens* of those cities. Notice that it does not appear to be the case that the timocratic man at 549c2-6 is being viewed as a citizen of a timocratic state; rather he is being viewed as a father in any arbitrary city that is not well run, who turns his back on the offices and honors of the city, and whose wife and servants all influence the son behind the father's back. Plato makes no point about wives and servants in the timocratic city. It is true that Williams's passage 544d6-e2 says that there are not going to be *more* kinds of constitutions than there are kinds of individual psyches, because in the loose sort of way identified at 435e-436a, constitutions of the cities come from the kinds of individuals in them.

Now it is not that Williams is unaware of the sorts of points I have been making against each of the propositions and models he extracts from his readings of particular passages in Plato. But in a diagnostic stance all too typical of analytical philosophers, he does not hesitate to attribute *all* of them to Plato, yielding an incoherent mess and attributing that too to Plato – instead of regarding this as a sign that he is not grasping Plato's whole picture. (Compare the argument of Sachs discussed in n.27 below).

I do not attribute to Plato any of the preposterous principles Williams attributes to him – not the simple-minded semantical principle discussed in n.21 above, not the part/whole rule, and not the 'predominant section' rule, which gives Williams the position that Plato is committed to the view that as a democratic city is a city with very various character, so each democratic citizen is a citizen with very various character. (As one might infer from the apostles being 12 that each apostle is 12). Williams gets this last conclusion by attributing to Plato the inference from (JCITY) to (JCITIZEN), which, I have argued above, Plato is committed to denying, and *would* deny.

²³ Socrates constantly injects into his arguments, first with Polemarchus and then with Thrasymachus, the suggestion that justice, like medicine, navigation, and the like, is a science or expertise useful for gaining some end. Neither Polemarchus nor Thrasymachus start off as if they think justice is an expertise or science. Thus Polemarchus, like Cephalus, seems to start off with the idea that justice is *doing* certain things (telling the truth, returning what has been deposited with one, rendering to each what one owes, rendering to each what is fitting) or, at any rate, being the sort of person who does these things; and Thrasymachus seems to start off with the positivist idea that justice is nothing more than the oppressed doing certain things which have been commanded by the oppressors and as a result almost always 'getting less.'

For Socrates to force either of these initial accounts into the mold of expertises or sciences might thus appear inexplicable – unless Socrates himself endorses something like this idea that justice is an expertise. But Socrates does precisely force these two initial accounts into the mold of expertises or sciences. Thus (a) Socrates' questions force Polemarchus to the view that justice is an actual *expertise* at helping friends and harming enemies. (And apparently the problem with this view is not the *expertise* part, but the *harming enemies* part: 334c-335e). And his questions force Thrasymachus to the view that justice is an *expertise* at getting one's

own advantage (an expertise at getting the better of others) by ruling others – an expertise had by only *some* of those oppressors who when in power declare certain things just. (The problem Socrates raises about this is not with the *expertise* part, *nor is it* with the *advantage* part – on which see n.45 below. The problem lies rather with the belief that one *could* gain one's own advantage by getting the better of – and thus harming – those under one's care: 345b-347e). This appeal to expertise in fact forces Thrasymachus off of his original positivistic account which appeals only to what *any* oppressor declares just.

By what right does Socrates *read into* these views of his interlocutors an assumption about expertises so unnatural to the accounts they first offer, and in terms of which they are presumably throwing down their challenge to Socrates? On standard modern views of what people *mean* or *have in mind*, if we were to go by what Polemarchus or Thrasymachus had in mind or meant when they put forward their views, we would not expect to get that they had in mind, or meant, some form of expertise. It is true that moderns might be prepared to accept that Socrates actually helps Thrasymachus *out*, by converting the justice positivistically imposed by the rulers (which could at best be an *attempted* exploitation of the weaker) into an actual *science* of exploiting the weaker. But by Sachs's lights, this still involves a change of subject from the rules actually laid down. (It is interesting to note that when Plato comes 'to the interest of the stronger' in the *Laws*, at III.714b-d, he returns to the positivistic version exploded in *Republic* I. There is no sign here in the *Laws* of the ruler *qua* ruler who makes no mistakes in exploiting others).

So shouldn't Sachs also suppose that Socrates has changed the subject on Polemarchus and Thrasymachus – from a certain sort of character, or a tendency to do certain sorts of deeds, to an expertise? (More on justice as skill or expertise in the next two notes).

²⁴ An expertise pure and simple is one that admits of the ancient axiom that there is the same science of opposites (see, e.g., Aristotle, *Metaph.* XIII.4.1078b26-7, *NE* V.1.1129a11-17, VI.5.1140b20-25). We might call it the Axiom of the Ambivalence of arts, sciences, or expertises. This gives us, in the *Hippias Minor*, that the person good at running fast will also be the expert at running slowly, while the inexpert person is the person who will err *unwillingly* at running fast. It will be the person expert at running fast who will err willingly at running fast.

Such a conception of expertise pure and simple has seemed to many to show the *Hippias Minor* actually *reducing to absurdity* the notion that justice is a science (expertise) or power (375d7ff). For if justice is a science or power, it will make the just person be that person – 'if there *is* indeed such a person' (376a4-6) – who does unjust deeds willingly. This reading of the *Hippias Minor* as reducing to absurdity the suggestion that justice is a science of any kind (including, e.g., the science of goods and bads) seems to me clearly mistaken. The point of the *Hippias Minor*, I argue in my 1973b, is that ambivalence does *not* reduce to absurdity the idea that justice is an expertise pure and simple, since in the case of *that expertise alone* no one errs willingly at it (see the escape clause in quotes above, which Hippias entirely misses – as some modern readers of the dialogue seem reluctant to recognize its significance). I take it that it would be bad dialogue-writing on Plato's part *both* to argue that ambivalence shoots down the idea that justice is an expertise pure and simple, and *also* to argue that there is no difficulty caused by ambivalence in the case of this one expertise pure and simple. For contrary views, see Kraut 1984, 311-316, Kahn 1995, 113-121. (Of course to say that the sciences are ambivalent is not to deny that the *function* of the science of medicine is healing and not the opposite).

²⁵ A question may be raised here as to whether the talk in Book I of 'ruling oneself' at 351b7ff, esp. 351e6-7, 352a5-8 with 442d1, suggests more the notions of justice as requiring strength of *character* than the idea of justice as an expertise pure and simple. (Compare also *Gorgias* 491d: also interestingly enough in the context of a discussion of justice vs *pleonexia*). Such a departure would embrace already the 'parts of the soul' doctrine which brings with it the idea that a virtue requires, over and above expertise pure and simple, elements of character-training and even a certain sort of nature. Like Aristotle's *phronēsis*, wisdom in the *Republic* (after Book I) cannot be acquired except by someone who has had the right training of the irrational parts of the soul, and the right habituation. For Plato and Aristotle, it is impossible to gain wisdom by simply aiming at

intellectually getting things right – by simply aiming at succeeding in Socratic dialectic. It is true that there is no wisdom *without* intellectually getting things right and without succeeding at Socratic dialectic. It's just that the wrong nature or character destroys (by non-intellectual means) any chance of success in those intellectual endeavors. (On the need for constant testing to make sure the relevant nature and character is present, see, even within Books VI-VII, esp. 502e-503a, 503e-504a, 537aff, 540a; and on the relevant nature and character, see 485a-487a, 489e-490d, 491a-492b, 494a-d, 518e-519b, 522a [on good habits without knowledge], 527c, 539d. Books VI and VII are *not* just metaphysical digressions from the enterprise of Books II-IV). It is probably because of this presence of character in all those with wisdom that Ambivalence disappears from the rest of the *Republic*.

To return to our passage in *Republic* I, if such talk of ruling oneself brings with it the idea, first announced in *Republic* IV, that

DESIRE/ACT an irrational desire may cause one to act contrary to one's desire for what is best (in the manner of Aristotelian *akrasia*),

then this passage in Book I would involve a conception of justice as requiring character rather than a conception of justice as an expertise pure and simple. For then, Book I would bring with it a psychology of action inconsistent with the Socratic view that all action proceeds from one's desire for what is best. And indeed I have until recently held that such talk of self-mastery almost certainly brings with it something like the anti-Socratic 'parts of the soul' doctrine. But ongoing discussions with Vasilis Politis have led me to see the possibility of denying this connection in the *Gorgias* and *Republic* I: see the remarks towards the end of the first para. of n. 23. (I do not know whether Politis would agree with my present position). In any case, even if *Republic* 351e-352a does presuppose the anti-Socratic position of *Republic* IV, it would remain the case that all of the other arguments of *Republic* I involve the Socratic intellectualist treatment of virtue as a science or expertise pure and simple.

²⁶ The assumption here is that whether or not Thrasymachus is right in what he is claiming, he does know *what he is claiming*. Put otherwise, whether or not he knows with certainty precisely *which* individual acts are just, he must know with certainty what would *count* as a just act. He must know what he *means*. Here we can see surfacing in Sachs's position the notion of meaning as knowing 'what would count as' – a central feature of meanings (senses) in Frege. So too Prichard 1928, in 1968, 207, 214-5 (who speaks of us as knowing with certainty what kinds of acts are just), and Geach (n. 57 below). On Frege's notion of 'thought' or 'proposition,' which exemplifies particularly clearly the idea of 'what would count as,' see further n. 32 below.

²⁷ Actually, Sachs holds that at *Republic* IV.442d10-443a11, Plato *does* make a relevant attempt to justify his new conception of justice by the criterion of non-performance of acts of certain kinds (see the first three paras. of the section entitled 'THE FALLACY'). And it must certainly be conceded to Sachs that within this passage, Plato is correcting any impression he may have given earlier that his account of justice will lead to approval of acts of adultery, sacrilege, stealing, etc. Just so, Plato says earlier (433e3-434d1), in a similar bit of reassurance, that his idea of the justice of the city will lead amongst other things to the preserving of property. But for the rest, so far is it from being the case in 442d-443a that Plato is judging psychological well-adjustment against the criterion of non-performance of certain kinds of acts, that at 443b7-444a2, Socrates says that we judge *acts* to be just or unjust according as they create and preserve the relevant well-adjustment of the soul. Not conversely, as Sachs would have it. (Actually, Sachs has it both ways. The earlier passage, he says, shows Plato granting that his conception of justice must conform to the ordinary conception of justice. The later passage, Sachs says, shows Plato abandoning the ordinary conception for the Platonic conception. Sachs's impulse here is not dissimilar to that of Williams on which I comment in the 2nd last para. of n. 22. Against Sachs's move here, I suggest that the passage, both in its earlier and in its later parts, is surely to be taken as a whole. And so taken, it is saying we judge whether an act is just in terms of whether or not it creates or preserves well-adjustment, and not conversely). It is true that, to make his account of justice *prima facie* plausible – something any moral reformer must do – Plato wants to say, 'There is no reason to think this new account of justice will

lead to theft, murder, sacrilege, and so forth.’ But that is not to say that Plato judges justice by this criterion; and nothing stops Plato from departing from some ‘moral rules’ either. [In any case, Sachs admits, in the last paragraph of his article (with his note *ad loc.*) – as if as an afterthought – that as the Ph.D. dissertation of G. Santas has shown him, Plato recognizes no exceptionless moral rules. At this point, we may ask: How can Sachs admit that Plato thinks there are no exceptionless moral rules and also have Plato admitting that there is a vulgar conception of justice, consisting in ‘the non-performance of acts of certain kinds’ which he, Plato, is conceding any account of justice must conform to? Why does Sachs not complain of Plato’s inconsistency here? Sachs seems to me simply to have *read into the text* this admission of a vulgar ‘criterion’ for justice].

²⁸ Santas 2003 has objected to the 2001 version of this paper that in the text of Book I, we find no reference by Thrasymachus to the real nature of justice, but only to such things as the advantage of the stronger, of which I imply in the main body of the paper that we may say that it represents *not* the real nature of justice, but what Thrasymachus *thinks* justice is. So, he says, there is no evidence that Thrasymachus thinks he is speaking of the real nature of justice.

The issue here is the identity of Thrasymachus’ conception. Santas and I agree, I suppose, that aside from the question how Thrasymachus conceives of justice, the advantage of the stronger is not the real nature of justice. We also agree that how Thrasymachus conceives of justice is as the advantage of the stronger. But that is not the end of the matter. The question is what *else* there is to Thrasymachus’ conception. Does he conceive of justice as

The advantage of the stronger *tout court* (regardless of what else may be true of it)

or does he conceive of it in the following way:

The advantage of the stronger, i.e., the real nature of justice?

Does Thrasymachus conceive of the advantage of the stronger as (in some no doubt deflationary way) the real nature of justice? Santas says No, I say Yes. My idea is that what Thrasymachus intends to refer to, as viewed by him, is

the real nature of justice, whatever it really is even if it differs from what I think it is – and if I get it even slightly wrong, I want you to fix it up – but which I can assure you is nothing but the advantage of the stronger.

Now my problem with Santas’s objection is that it seems to me to invite us to choose one of the *following* as representing Thrasymachus’ conception:

a The advantage of the stronger *whatever the real nature of justice may be – and I don’t care what else it is,*

as opposed to

b The real nature of justice *which is something different from the advantage of the stronger.*

(which Santas seems to suppose is the conception I am attributing to Thrasymachus). But while the choice of (a) represents the approach characteristically taken by analytical philosophers – going for whatever reference is determined by the meanings of the word used *independently of what else may be true of the actual reference* (see, e.g., n.30 below) – it is the approach I am contending against throughout this paper. In fact, as I see it, the issue is *not* to choose between (a) and (b) – not initially, anyway. That particular choice could only arise at a later stage of the argument, once we had convinced Thrasymachus that he is wrong to identify the real nature of justice with the advantage of the stronger. So it is not a choice we are faced with when we examine his

argument with Socrates. How Thrasymachus thinks of what he is talking about in the course of the discussion – and hence what Thrasymachus intends to refer when he uses the word ‘justice’ during that discussion – is neither in terms of (a) nor in terms of (b), but rather (as suggested above) in terms of

- c That real nature of justice whatever it really is (if I am in error in some way, you fix it up), i.e., the advantage of the stronger.

To say this is to say that there is an incoherence in Thrasymachus’ intention to refer. The ‘i.e.’ will eventually, as it were, ‘come apart’ on Thrasymachus – he will not be able to have both what is on the left-hand side of the ‘i.e.’ and what is on the right-hand side. That being so, if we are to ask whether (c) shows that what Thrasymachus was intending to refer to was the real nature of justice – this is the issue that divides Santas and myself – then we must ask what Thrasymachus *would* say he intended to refer to when we show him that the ‘i.e.’ has indeed ‘come apart.’ Obviously he is not going to insist on sticking with the ‘i.e.’ that has ‘come apart’ in (c). So it is at this point where the choice between (a) and (b) first arises – when we are considering the second choices, or retrenchments from error, that are available to Thrasymachus once he sees his error.

At the point of seeing his error, (b) would surely be the correct answer to the question what Thrasymachus was intending to refer to when he used the word ‘justice.’ (b) is not, of course, the answer to the question what Thrasymachus’ *first* answer would be to the question ‘What is justice?’, but the answer to the fall-back, second question he must answer when he discovers that the ‘i.e.’ has ‘come apart’ on him. That is, I am suggesting that the only good answer Thrasymachus could make here is exactly what I have claimed. Put in another way, what Thrasymachus is saying is, in effect:

I *did* intend precisely to speak of the real nature of justice. It’s *that* of which I was saying that it didn’t make people happy. But it appears I was under a misapprehension about what that real nature was.

In other words, it *was* Thrasymachus’ intention to speak of the real nature of justice – throughout the discussion. (In Penner 1991, I have made a parallel claim about what Helen would say was what she had wanted when she made the mistake of running away with Paris). This is to say that within (c), it is the real nature of justice that ‘dominates’ and not the advantage of the stronger. And this, I say, establishes that the disagreement between Socrates and Thrasymachus is about the real nature of justice, as I claimed. My point here is basically the same as that I make to a slightly different objection in sec.VII below.

Thus I do not find Santas’s objection compelling. That said, I need to say how enormous is my debt to my continuing dialogue with Santas’s many distinguished writings on Socrates and Plato and with Santas himself. There is no other Plato scholar over the past half-century or so to whom I owe more.

²⁹ This notion can be seen also in the crucial (and enormously consequential) remark at Frege’s *Grundlagen der Arithmetik*, sec 47: ‘... to decide what [a given sentence] is about, we do not need to know whether it is true or not, nor for what reasons we believe it to be true.’ (The argument here is meant to establish that ‘All whales are mammals’ is not about whales but about the concept *whale*, and, correspondingly that existence is not a property of the things said to exist, even when such things *do* exist. The arguments of Hume and Kant on existence not being a predicate are no better, and perhaps a tad less clear). It is characteristic of philosophers to insist that we do not need to know the truth value of a proposition before saying what it is about. Evidently the theory of what a proposition is, is uniform with that described in the preceding note. Compare also the position of Geach in n. 57 below.

³⁰ Here, briefly, is how I understand the ideas of meaning and reference. Consider the two expressions ‘The Morning Star’ and ‘The Evening Star.’ We now know, though people did not originally, that these two expressions stand for the same object in the world: that is, they have the same *reference*. But according to believers in meanings, these two expressions have different *meanings*. The idea here (roughly) is that these two meanings are given to us in the dictionary, as two different sets of instructions for finding the reference in the

world – two different conditions or criteria that a thing in the world must satisfy if it is to be the reference. Thus the sets of directions we find in the dictionary are the meanings. The one set of directions instructs us as to where to look in the sky in the morning, the other instructs us as to where to look in the sky in the evening. One reference, two meanings. The conclusion is obvious: meaning is not identical with reference.

Now, it is commonly held by believers in meanings that, although meaning cannot (for the reasons just given) be *identified* with reference, nevertheless meaning *determines* reference: that is reference is what you find when you (correctly) apply the meaning (the set of instructions for *finding* the reference) to the world. This idea that meaning determines reference may be held in the form just described, and also in considerably liberalized theories of what a *meaning* is and what *determining* is (e.g., for indexicals – in modern philosophy, words like ‘now’ ‘there,’ and the tense of verbs) meaning plus ‘*context*’ (understood here, somewhat narrowly, in terms of those features of the world that settle the reference of indexicals) applied to the world yields reference. But suppose that *no* such theory of meanings was adequate to the determining of reference. Then I claim that meanings that don’t in any acceptable way determine reference *don’t have any real philosophical use*. (That was the whole *raison d’être* of meanings.) Since I don’t myself believe that meanings in any form determine reference, I don’t believe in meanings. (It follows, of course, that I also believe that interpreters of Plato should not appeal to meanings in attempting to interpret the text of Plato). I realize, of course, that this is a highly controversial view, even after Quine and Davidson (who do however, ruinously in my view, hold onto a notion of *holistic* meaning). Can we do without the idea of meaning in one of these ways determining reference?

Let me indicate briefly, using a varied set of counter-examples, why I reject the view that meaning determines reference. ‘The man in the corner drinking a martini’ (Donnellan) may refer to the secret teetotaler in the corner drinking water from a martini glass. ‘My lawful wedded wife’ used by Oedipus may refer to Jocasta (who, being his mother, cannot be lawfully wedded to him). And as Plato tells us (*Cratylus* 387a), when we want to cut, we want to cut, not in accordance with what we *mean* by ‘cut’ (our conventions for ‘cut:’ 385d2-6 with 384d1-2, 385a1-4), nor in accordance with what we *believe* about cutting, but in accordance with the real nature of cutting (actions being done *kata tēn hautōn ... phusin*), even – I would add – if that is different from what is picked out by what we conventionally *mean* by ‘cut.’ Again, if a pre-Hippocratic Greek were to refer to *that Russian novelist in the corner with the Sacred Disease*, he would still be referring to Dostoevsky, even though the meaning of ‘the Sacred disease’ would require the novelist actually to be possessed by devils if the expression were to have a reference. But since there *are* no devils, no Russian novelist was ever possessed by devils. So if meaning determined reference, pre-Hippocratic Greeks would not in this case be referring to Dostoevsky. And it’s no good suggesting that pre-Hippocratic Greeks used an expression that we can simply enough correct for them. For what if it had never been discovered that there are no devils? Thus I do not believe that we should be impressed by theories that say the meanings of *corrected* expressions determine reference. We refer to cancer and Alzheimer’s all the time without any scientists being at all clear what they really are. Similar considerations apply to the first three counter-examples.

My (Platonic) treatment of these four counter-examples is not, I concede, the only way of responding to them. For example, as noted already, a more sophisticated proponent of meaning will appeal to indexical words like ‘this,’ ‘now,’ and so forth, and argue that it is not meaning, but meaning *plus* ‘*context*,’ that determines reference. This device can be used to make all the examples above into implicit indexicals. For example, treat ‘The man in the corner drinking a martini’ as *that* man in the corner drinking a martini. (The indexical ‘that’ addresses you to the present context, telling you to plug in the man *over there* in the corner drinking a martini. The indexical is a kind of blank check, which we fill in with the context). But now we must add the stipulation that context may *correct* the meaning being appealed to. (He’s not drinking a martini, but water from a martini glass. The problem is solved by saying that the indexical ‘that’ plus the context makes it clear it *is* the man *over there* who is being referred to. The indexical ‘that’ overrides the meaning of the rest of the phrase, namely ‘man in the corner drinking a martini’).

I take it that these more sophisticated proponents of meaning will insist that although we are in this case *correcting* the meaning of ‘the man in the corner drinking a martini,’ we are nevertheless not *abandoning*

meanings. But with what plausibility? It will be said that the meaning of ‘contextual’ words like ‘that’ (‘contextual words’ being now restricted to indexicals) are nevertheless still relevant to determining the reference, since ‘this’ and ‘that’ will determine the relevant ‘*contexts*’ in accordance with *their* meanings. Thus, you could not be referring to *that* man in the corner if you said ‘*This* man in the corner.’ So you’re still using meanings to determine reference, since you are still using the meaning of ‘that’ (plus the context) to override the meaning of ‘man in the corner drinking a martini.’ But why on earth would we not, in appropriate circumstances, also override the meaning of ‘this’ in such a case? Why wouldn’t a hearer in the know say to himself or herself, ‘Oh, Jones is talking about *that* man in the corner who’s drinking water from a martini glass.’ As I see it, this last sort of example shows that meanings (meanings of indexicals, supposing that there are meanings at all) do not determine even how ‘*context*’ figures into the determining of reference.

Now, either all expressions are implicitly indexical, or some are not. Case I: If some expressions are non-indexical, then the meaning of those expressions will once more, by the arguments above, not determine reference (nor will the meanings of indexicals even determine how context enters into determining reference). Case II: If *all* expressions are implicitly indexical, and meanings may always be overridden, including the meanings of indexical words, it is hard to see how we could preserve the notion of meaning as always (in one way or other) determining reference. But, as I have already remarked above, if meanings don’t determine reference, it is hard to see what reason there could be for believing in them at all. For that was their entire *raison d’être*: to determine reference. In Case II, of course, one is in effect saying, in my terms, that ‘context’ *includes* the reference intended by the speaker(!). (Why wouldn’t it? my friend Larry Levy wondered to me). In that usage we get, once more, that reference would actually trump what meaning determines. Once again, meanings cease to play any theoretical role. I am myself completely happy with the view that context includes the speaker’s intended reference. But if we say this, then there are two other things that need to be noted: First, this understanding conflicts with modern treatments of ‘*context*’; and, second, this will have the effect that context is now not merely *supplementing* meanings in determining reference, but *supplanting* them altogether.

It might be thought that even if meanings do not determine reference, meanings could still have a function with respect to problems about the analysis of people’s beliefs – including the problems caused by Frege’s paradox of identity: How is it that some can believe that the Morning Star is a planet, yet not believe that the Evening Star is a planet? The idea here is to analyze *what is believed* not in terms of the *reference* of ‘The Morning Star’ and ‘Evening Star’ – since then the two beliefs would be the same – but in terms of the different dictionary directions for finding the reference which are implicit in the two different beliefs. (Since James doesn’t always know the truth about the *reference* involved, his belief that the Evening Star is not a planet may be false; but since he implicitly knows the *meaning* of ‘The Evening Star,’ he knows *what* he believes. See n.26 and text thereto, on the claim that Thrasymachus must know what challenge he is making). But promising as this Fregean move is, it must fail if meaning does not determine reference. For *what I believe* when I say ‘The man in the corner drinking a martini is a spy’ cannot be given by *those* dictionary instructions for finding the person I am referring to. So whatever the solution to the paradox of identity, it must be other than this one. (I explore the larger question about belief-contexts in my still unpublished *Plato and the Philosophers of Language*).

Thus I do not myself accept the view that meaning determines reference – or even that meaning plus context (as understood by moderns) determines reference. But it *is* that view that meaning determines reference that generates the Fregean theory of questions which I am saying Sachs is in effect relying on in his charge that Socrates changes the subject on Thrasymachus. There is more on meanings in nn. 32 (sentence meanings and propositions), 33, 36, and 37, below.

³¹ It is true, of course, that believers in meanings can *say* that the disagreement is ultimately about reference, because it is about whatever reference is determined by the meaning. Alternatively, they can admit that *de dicto* the disagreement is about meaning, as Frege says, but that does not stop anyone from saying that *de re* it is about the reference determined by the meaning. (Compare the designatory function *des* (σ , r) from senses to references in Kaplan 1969). But here the determination of reference by meaning, or the designation of reference by sense, occurs in an entirely extra-mental way. The person is not (except in this extra-mental way) thinking of

the reference. So what they *have in mind* in the disagreement – what they agree they are disagreeing about – is not the reference. In any case, if meaning does not determine reference, these suggestions will not work.

I myself believe these difficulties expose a serious difficulty in Fregean theories of reference, and therefore also in the kind of methodology that guides Sachs in his article. How could disagreement fail to be about the reference? Of course one consequence of that will be that we don't in general *know* what the thoughts are over which we are disagreeing, nor do we *know* what the reference of 'justice' is over which we are disagreeing. I present a new, anti-Fregean, theory of reference in my *Plato and the Philosophers of Language*.

³² The Fregean notion of 'thought,' ancestor of the modern notion of 'proposition,' is (indexicals aside) the meaning of a sentence. This tells us *what would be the case if the sentence were true*. So we may consider whether some sentence is true without knowing whether or not it *is* true. We do so by knowing what state of the world *would count* as making the sentence true, even if we don't know whether or not that state exists. (The truth-conditions of the sentence). This we know by knowing the meaning of the sentence, for that meaning is precisely what tells us what state of the world *would count* as making the sentence true. If I know that Bill Clinton is the 42nd president, but Peter does not, then what Peter believes when he believes that Bill Clinton is a former Rhodes scholar is different from what he denies when he says that the 42nd president is *not* a former Rhodes scholar. For the meanings of the two relevant sentences are different. Peter knows what he means, and plainly, since he affirms the one sentence and denies the other, he means something different by the two sentences. (In fact, as Evans points out in his posthumous 1982, 18-19 – he seems to attribute the point to Paul Benacerraf – because they are different beliefs for Peter, they must also be different beliefs for *me*. But this feature of Frege's views does not concern us now). This is because Frege supposes – it is a condition of his treatment of the paradox of identity (n. 31 above) – that whatever my ignorance of the reference of my words, I must know the *meaning* of my words. By contrast with this appeal to propositions, I ally myself with the Platonic line of thought laid out in sec. VII. There is much more on the deleterious effects of the notion of proposition on analytical philosophers interpreting Plato in my C.J. de Vogel address on the so-called 'Socratic Elenchus' to the 2004 meetings of the International Plato Society in Würzburg.

³³ These Fregean underpinnings of Sachs's argument are identical with those which also issued from Ithaca NY at about the same time, in Norman Malcolm's great (but perfectly astonishing) book *Dreaming* (1959). Here Malcolm argued that proponents of the use of Rapid Eye Movements to show that dreams are processes that take place during sleep *are not talking about dreaming as we understand it*, since that, after all, is given by the *meaning* of the word 'dream,' which meaning could allow 'having dreamt' only to be a tendency to tell a story upon awakening. (How could you ever teach someone to use the words 'I am now dreaming,' except in terms of telling stories upon awakening?) Proponents of REMs as indicators of dreams would be committing (what Sachs calls) a fallacy of irrelevance in saying that since dreaming in the Dement-and-Kleitman sense is a process that takes place during sleep, therefore dreaming in the ordinary sense (or vulgar conception) is a process that takes place during sleep. As I pointed out in my (1973b), 144-5, it is in this way that, in the view of philosophers like Malcolm, analytical philosophy gains a conditional stranglehold on the future of science.

³⁴ Putnam's 'division of linguistic labor' (see note after next) is from this point of view a failed attempt to capture the 'blank check' idea – failed, because even the scientific community is often wrong, and must in any case issue blank checks.

³⁵ I am grateful to Andrea Veltman for pointing out the importance of my addressing the present objection. The objector's point here is simply a more pugnacious variant of the Santas objection discussed in n. 28. Where my preferred (c) in n. 28 uses the useful expression 'i.e.,' my preferred (10c) here uses an identity.

³⁶ The example is from Putnam 1962, 220-221 with 218-219, 224, 235, a discussion of Malcolm 1959 (n. 33 above), though the lesson I draw from the use of the example differs from the one Putnam draws. I want to say that – at least for this example: I am not saying, any more than Putnam is, that all cases in the history of science work out so well – the reference of 'acid' remains the same throughout the centuries, and that when scientists

disagree with each other they disagree about the reference. But if there *are* meanings, then the scientists over the centuries will have been assigning different meanings to ‘acid,’ and so, by Sachs’s lights, will not have been disagreeing with each other. By contrast with Sachs, Putnam wants the *result* I get, that the reference of ‘acid’ stays the same over the centuries. But he wants to get that result without giving up meanings. Thus he wants to say that ‘acid’ and ‘dreaming’ are special kinds of expressions, called ‘natural-kind terms’ and that these natural-kind terms *do* have meanings. Natural-kind terms *mean*, roughly, *whatever natural kind it is that our term in fact picks out*. In Putnam’s usage, a natural-kind term is a sort of indexical – it involves a reference to the natural kind involved in *this* case (or at any rate the natural kind involved in *this kind* of case). Looked at in another way, it is a variant of what I have been calling a ‘blank check’ theory: para. 3 of sec. II above, as well as n. 30 above. But it is a ‘blank check’ theory where the blank check is built into the *meaning* of the term. ‘Acid’ *means* a certain natural kind, explanatory of roughly the following extension: [Here we specify the extension currently assigned to ‘acid’] In later works, Putnam spells out the kind of blank check involved in terms of his division of labor theory of meaning, glossing the natural kind that is in fact picked out in terms of *what scientists would say, or will eventually say, it is*. (This seems to me to be not quite the right kind of blank check: for what if scientists will *never* figure out what certain key scientific entities are, or if they will *think* they have, but have not? Nonetheless, the turn to blank checks is common to Putnam’s theory and the purely referential theory I espouse).

The situation, then, is this: Putnam thinks that this meaning of natural kind terms (‘whatever the natural kind may be in cases like this’) does determine its reference – by giving a blank check to whatever scientists declare, or will declare, the natural kind in fact to be. My disagreement with Putnam will be clear if I say that in my view we might as well issue the blank check to whatever the natural kind may really be, instead of issuing it to whatever the experts currently say or will in the future say. But then why not drop the meaning altogether and give a blank check to the reference, saying that in the case of natural kind terms, there is no meaning *additional to the reference*? Such is the view in Stampe, 1972: that natural-kind words don’t *have* a meaning: all they have is a reference – a causal origin, Stampe supposes. Leaving aside for now questions of causation, what I want to do is to extend Stampe’s view that natural-kind words don’t have *any* meaning to all other words – verbs, adjectives, adverbs, prepositions, conjunctions, etc. Thus my proposal – it is also the view I attribute to Plato – may be viewed as the proposal, starting with natural-kind words, to understand *all* words in terms of their *reference* and to drop the notion of meaning altogether. (The theory of reference is where Quine and Davidson started out: unfortunately, they circled back to another kind of meaning, *holistic* meaning, which is in turn is understood in terms of derivability from the Tarskian truth-definition for the language). On the view I espouse, the reference of ‘if,’ for example, is given not by how we use those words, but by what the real nature of supposition is. (For example, is supposition – as I believe – what some Stoics, some medievals, and Frege say it is, namely, material implication? Or is it, as relevance logicians, and proponents of modal logic claim, something involving relevance or alternative possible worlds?).

³⁷ Plainly, I am also anxious to reject the view of more recent interpreters who in one degree or other have followed Sachs’s line. Irwin 1995, 385 even points out that Sachs’s criticism is anticipated in Grote (on whom see next note). Of later writers, Annas 1981, 213 formulates Sachs’s accusation in terms of *concepts* and *intuitions*. She says,

Thrasymachus’ challenge arose from comprehensible doubts about the worth of justice as ordinarily pursued. So Plato’s answer must keep in touch with the original concept.... ..Plato sees the need to let his account be constrained by our central intuitions about the ordinary concept; otherwise he will have missed the point.

(See also 164-7 for more on Annas’s belief that Plato ‘sees the need to let his account be constrained’ by our ordinary concept). In taking Plato as trying to conform to ‘the original concept,’ Annas is evidently following Sachs’s line.

Irwin 1995 is altogether more wary of meanings and concepts than Annas. (See, for example, 49, 366n.5, 69-

70). On the other hand his notion of an Epictetan ‘articulation of our preconceptions’ (49-50 with 48) represents a liberalized notion of meaning rather in the manner of Putnam (preceding note). Even within such a liberalized notion of meaning, Irwin will argue that Socrates

cannot suppose that people are always or usually mistaken as they are in these particular cases; evidently his inquiries could not get started unless he thought his interlocutors could recognize *some* examples of the virtues fairly reliably. (50, my italics).

In my view, there should be no bar to rejecting all examples whatever, along with meanings, and going for whatever the real thing is that is the reference – even if the reference is quite other than what any meaning of ours, however liberalized, would determine. (Here I am in agreement with causal theories: Why do we have to know in advance that at least some of the standard examples are correct?) The problems Irwin sets up for Plato and tries so strenuously to solve are in my view generated by a false theory – the theory of meanings. Which is not to say that as Irwin makes his way through these (as I see them) pseudo-difficulties, he does not say a number of things with which I strongly agree.

Having said this, I should confess that I now disapprove of Penner 1973b, 145-7, for the suspicion it gives (even if it is part of an *argumentum ad hominem*) of arguing from liberalized meanings (in effect, Putnam’s cluster concepts). Fortunately, nothing in the rest of the article involves any such notion.

None of the above is to deny that Sachs’s charge must work against any reading of Plato that we find in the work of interpreters who happen to believe in propositions, meanings, and the like. If you believe in meanings, or if you think that before you can even look for an answer, you must know what the question is, then you must face Sachs’s challenge.

³⁸ This is, as it were, the old accusation of changing the subject, and is made without any appeal, or at any rate any need for appeal, to meanings, concepts, conceptions, or linguistic intuitions. The point is simply that what Plato talks about *doesn’t seem to be the reference* of the word ‘justice.’ If we call this ‘the old accusation,’ Sachs’s decidedly 1950’s accusation is, as it were, ‘the new accusation.’

Grote 1888, ch. 34, also speaks in terms of a Platonic sense vs. an ordinary sense. But the gravamen of his charge is not so much the new Sachsian charge that Plato commits a fallacy in reasoning, as the old charge that Plato’s account of the *reference* of justice’ is wrong. Justice cannot be what the *Republic* says it is – a single internal ‘entirely self-regarding’ state (‘a state of internal happiness to the agent’). What it has to be, Grote insists, is rather a state of exchange – ‘reciprocity’ – between its possessor and others, in which the possessor exchanges the fulfilling of his or her obligations to others for the receiving of his or her rights from others. (In just this way, Grote will have thought the Greatest Happiness of the Greatest Number will have been served – the true idea of justice, in Grote’s view). This should, indeed, have been the upshot, Grote says, of taking seriously the idea of justice being based naturally on our need that each ‘does his own’ (369bc with 370c) – presumably since each needs that the other does his own. It should also have led Socrates to grant that fathers *do* act correctly in praising justice to their sons for its consequences in how other people see us. Socrates is quite wrong to suggest (368c) that it would be a slander on justice to allow such consequences to be the real basis for justice being a virtue. (Grote usefully cites *Laws* 950b as an implicit criticism of *Republic* Book II). I need hardly add that what Grote’s argument requires is that Plato is interested in (JCITIZEN) not – as I have been claiming – (JINDIV).

³⁹ So I read *Republic* IV.438a1-4, which I maintain says that not all desire is for the good. The desires spoken of here I assume to be desires productive of people’s *actions*. (Without this additional assumption, of course, we will not yet have this passage contradicting the Socratic psychology of action. Thus one could have other desires such as thirst and hunger without its being the case that one acts on them without thinking that satisfying them would be the best thing to do and accordingly desiring to do that best thing to do. It is this desire to do what it is best to do that, according to Socratic parts of the corpus, is productive of people’s actions

– not their thirst or hunger, much though they be involved as *part* of the situation in which the decision is made to do what is best. Compare nn. 22, 25 above).

This new Platonic, anti-Socratic account of the psychology of action, I used to call ‘Psychological Irrationalism’ because of its view that irrational desires could out-muscle one’s desires for the good. But since Plato’s ethical theory still requires for virtue that one be wise (and so, by the parts of the soul doctrine, that one’s rational part have knowledge), I now think the name ‘Limited Psychological Rationalism’ more appropriate.

As against my assumption that the denial that all desire is for the good represents an affirmation that irrational desires from the two lower parts of the soul can cause one to act contrary to one’s desire for what is best, there are interpreters who claim Plato is here only speaking of irrational desires that oppose reason by (non-rationally) pressuring reason to change its estimate of what is best – and so which do not by themselves cause action. In effect, this suggestion has Plato affirming here not the possibility of Aristotelian *akrasia* but the possibility of Aristotelian *akolasia*, in which the agent still does aim at what he believes to be good: it’s just that irrational desires have corrupted his beliefs as to what the good is – though in Aristotle it is not so much the belief as the desire for good that is corrupted). The case of Leontius, however (439e6-440a3), who does act on his irrational desires, seems to show this cannot be made out. I shall be discussing these matters elsewhere.

A different sort of rival to the present account of Plato’s doctrine here argues, in effect, that the ‘irrational’ desires of the two lower parts of the soul are not irrational at all, but rather quite as rational as Reason’s desires (quite as calculative of the best means to one’s end as Reason is, it being solely in their respective ends that the lower parts differ from rational part). Thus, on this account (see, for example, Irwin 1995, Bobonich 2002), the spirited part rationally maximizes honor, the appetitive part rationally maximizes bodily pleasures. This makes of Plato’s account of ‘rational’ vs. ‘irrational’ desires a kind of incommensurabilist rationalism which does not at all resemble the limited psychological rationalism of Aristotle’s theory of three sorts of desire. Thus for Aristotle, we have rational desire and two species of irrational desire: spirited and appetitive. (See, for example, *NE* I.13.1102b13-28, *De An.* III.10.433a25-6, 433b5-13). This rival theory in effect requires that Aristotle takes a seriously retrograde step from Plato’s position on ‘irrational desires.’ The view I have always taken, by contrast, is that Aristotle here has simply followed Plato – a view I think more historically plausible, especially given Aristotle’s frantic efforts to differentiate his position from Plato’s by saying he does not believe in the parts of the soul (*De An.* I.5.411b5-14). But this formidable incommensurabilist interpretation will need to be argued against on another occasion.

⁴⁰ I have presented and defended much of the supporting material on the Socratic psychology of action in my three O’Neil Memorial Lectures in the History of Philosophy in Albuquerque in March 2000 (to be published by the Department of Philosophy of the Department of Philosophy at the University of New Mexico), and in a paper delivered at the Pacific Division of the APA in April 2001, parts of which are presented in ch. 10 of Penner and Rowe (forthcoming). For supporting material, see my treatment of certain passages in the *Crito* in Penner Unpublished A, as well as my treatment of the parts of the soul doctrine in Penner 1990.

⁴¹ Kant’s ‘only thing good in itself’ is an unconscious parody of *Euthyd.* 278-282 – with quite different consequences. In the *Euthydemus*, it is because other things commonly called ‘goods’ are in fact only reliably good (only reliably lead to happiness: 278e3-279a5, 280b2 ff, esp. b7, d4-7, 279b2-4, 282a2, 4) if used in accordance with wisdom (knowledge) that knowledge is the only thing good in itself. This of course gets us precisely the opposite of what Kant wants – not a *categorical* good, but merely that *hypothetical* good of the sort Kant speaks of in his ‘counsels of prudence.’ See above, n. 14.

⁴² I have argued elsewhere (see n. 40) that *Apol.* 25c-26a shows Socrates arguing, in effect, that harming those around you can only be the result of ignorance of the fact that such harm to others harms you, so that punishment is *never* appropriate, only instruction. (It is true, as Ruth Saunders has pointed out to me, that there are occasions where punishment may *per accidens* give the person punished space to think, and so the space to

prepare the relevant self-instruction. But even in such cases, it is still the instruction, not the punishment, that addresses the ignorance; and it is the ignorance, not something punishment in general could plausibly be supposed to be directed towards – as on the ‘good swift kick’ theory of education – that needs to be dealt with).

This fact, that error is due to ignorance is, I claim, the correct conclusion for the Socratic determinist to draw. I have also argued that *Hipp. Min.* 372a-e show Socrates similarly rejecting the standard account of responsibility in favor of the intellectualist, determinist view that goes with ‘No one errs willingly.’

I shall not discuss here the way in which ‘no one errs willingly,’ when it shows up in Plato’s later dialogues no longer allows the inference to ‘all error is ignorance,’ some error being *madness* (*Tim.* 86b-87b) or disease (inner discord or conflict) (*Soph.* 227d-230) or just plain old *akrateia* or weakness of will (*Leg.* 734b-e, compare 680dff, 731b-732b). In fact, *Rep.* IV. 412e-414a shows that actions undertaken when Reason has a false belief are also unwilling. This passage also shows, incidentally, that those actions caused by irrational desires which Aristotle calls *voluntary* actions are, for Plato, *not* voluntary (= willing). (The treatment of punishment in *Laws* Book IX is just as deterministic as Socratic Intellectualism. It is the deterministic justification of punishment I have inveighed against in the main text just below). See also n. 48 below.

⁴³ That Sidgwick insists upon a categorical imperative as underlying all three of his methods of ethics – intuitionist, egoistic, and moral – is clear from 1907, 7-8, 31-35. Sidgwick is also very clear on the distinction between the sort of normative ethical egoism he embraces in his egoistic ‘method of ethics’ (and which he also attributes to Butler) and the sort of purely factual ethical theory he attributes to Ancient Greek moralists generally. (Though I do not agree with Sidgwick about Greek moralists generally, I do believe what he says is near to being right about Socrates and Plato: see n. 10 above). So let us now take his remarks (35-37) as applied to Socratic ethics. Sidgwick is here agreeing that Socratic ethics will involve no categorical imperative but simply a general desire for one’s own good or happiness, and so a purely hypothetical imperative. (‘Do this if you want to be happy, which of course you do.’) Socratic egoism will be what Foot has called a ‘system of hypothetical imperatives,’ though it would be better called a system of ‘counsels of prudence.’

The idea of the good involved here, Sidgwick allows (112), makes no appeal to anything but matters of fact, either actual or hypothetical, without any judgment of value or any ‘dictate of reason’ (categorical imperative). It is purely factual. This is just the picture I have been presenting in n. 16 above.

It may be wondered why Sidgwick does not take such forms of egoism as that in what I have been calling Socratic Ethics as a fourth method of ethics. The answer is that Sidgwick thinks (by a very curious argument) to be able to establish that hypothetical imperatives *themselves* presuppose a categorical imperative – a categorical imperative to act rationally (37-38), which makes it unreasonable to adopt an end and refuse to adopt the means indispensable to its attainment (37-8). Sidgwick sees well enough the reply – one which I myself would give, as I would also expect Socrates to give – that to adopt an end and refuse to adopt the indispensable means is not just unreasonable but impossible, since a decision not to take the means shows the end is no longer adopted. But Sidgwick – in effect embracing the possibility of *akrasia* – responds that people *do* sometimes refuse to adopt a means when they have not consciously retracted the end. In such cases, therefore, Sidgwick urges, ‘it can hardly be denied that we ‘ought’ (in consistency) to act otherwise than we do.’

This appeal to the fact that we *could have done otherwise* will, of course, not be usable against determinists of the sort I shall here maintain Socrates to be. Determinists, like Socrates (if I am right about him), and like myself, may therefore reject Sidgwick’s argument on this point alone. But even without determinism, I do not see why one would have to grant that the goodness of adopting the indispensable means in situations in which I have adopted the end is anything but a matter of fact. Not, at any rate, unless an opponent begs the question in favor of normativity by claiming that ‘goodness’ is itself a normative notion.

Sidgwick’s effort here, in his treatment of the hypothetical imperative, is to insinuate into the discussion the idea of ‘practical reason’ (37, 34-5) – a concept from which he hopes to generate the normative (the

categorically imperative). It is parallel to his use of a Principle of Rational Egoism within Practical Reason to get us from concern only for one's own *present* maximal good to one's maximal good over the rest of one's life, and, in parallel to this, a Principle of Rational Benevolence within Practical Reason which gets us from concern for one's own maximal good to the concern for everyone's maximal good. If Practical Reason can deliver this trans-temporal indifference, then why (Sidgwick wonders) should it not deliver a trans-personal indifference (381-4)? Hence Sidgwick thinks of himself as brought to utilitarianism. In a modern-dress version of this Sidgwickian argument, Nagel 1970 uses Practical Reason in almost exactly the same way to get us from one's own present maximal good to one's own maximal good over the rest of one's life, and from there to everyone's maximal good over the rest of their lives. Nagel even thinks (20-22) that the Lewis Carroll paradox can only be resolved by an appeal to Practical Reason, and so, apparently, that proponents of a system of counsels of prudence could not have a reason for detaching the consequent in *modus ponens*!) To such lengths will philosophers go in their concern to render the normative autonomous, and make the world safe for 'practical reason.'

For Socrates, all there is to this so-called 'practical reasoning' is theoretical reasoning applied to trying to find the best means to one's own maximal good. So why does Aristotle think that there is a special kind of reasoning called 'practical reason'? Here I can only gesture towards an answer (see also n.25 above). It lies in the same place as the answer to the question how *phronêsis* differs from the Socratic science of goods and bads. And the answer to that question is that Aristotelian *phronêsis*, like knowledge of the good in Plato's guardians *can only be achieved by proper habituation*, so that there are certain kinds of people – *akolastoi* (the wicked or self-indulgent) – whose ignorance is not reachable intellectually, but only via re-habituation, the *archai* of their rational parts having been destroyed (1140b11-21). This is the ignorance ('ignorance in the choice') that Aristotle holds – in strong contrast with Socrates – is not to be excused: 1110b24-33. I myself find the Aristotelian suggestion altogether odd, that absence of knowledge of particular circumstances (and presumably also absence of knowledge of certain technical and scientific disciplines) is to be excused, but lack of knowledge of the human good is not. This view is of course less odd if one adopts the Platonic/Aristotelian psychology of action, or if one is a legislator whose aim is to bring about certain sorts of behavior willy-nilly.

⁴⁴ The thesis that the fundamental desire causing all voluntary action is desire for the agent's own good and the thesis that a good person is a person good at maximizing his or her own good are, for me, two parts of what I see as the single purely factual theory which I call 'Socratic Ethics.' (On the purely factual, see the remarks on Sidgwick in the preceding note). Neither thesis is a matter of norms or imperatives or values.

The Socratic theory is thus not an ethical egoism *if* an ethical egoism must involve norms, imperatives, or values. It is, nevertheless, an *egoistic* theory in that the only fundamental desire involved is a desire for one's own good. If the theory is egoistic, the way it is egoistic is in its psychology. It is not an *ethical* (as opposed to *psychological*) egoism. From the point of view of the Socratic theory, the distinction between psychological and ethical egoism is not a useful distinction.

⁴⁵ See Price 1995, 16-17, where Price says of the apparent endorsement of hedonism in the famous discussion of *akrasia* at *Prot.* 352bff:

This is not the Socrates who resisted coercion to join in the arrest of Leon on the ground that he cared nothing about dying and everything about avoiding injustice (*Apol.* 32c4-d5, compare 28b5-8, d6-10), and refused to escape from prison on the ground that the prospect of death or any other misfortune counted not at all against the danger of acting unjustly (*Crito* 48c6-d5). It is true that the eschatology of the *Gorgias* would permit a hedonistic grounding of the crucial maxim 'It is worse to do than to suffer wrong' (e.g., 473a5), but scorning to let prudence weigh against morality, Socrates cannot wish to rest morality upon prudence.

Although Price allows here that the *Gorgias* admits of a prudential (egoist) justification of preferring death to injustice (not to mention what the *Protagoras* would admit of), it is in his view quite obvious that the *Apology*

and *Crito* could not possibly be offering a prudential (egoistic) argument for preferring death to the doing of injustice. Similarly Vlastos 1991, ch. 7 (the high point of the book according to Kahn 1992) takes it for granted that the point of the *Crito* on not doing harm back is the moral principle (11m), instead of my egoistic truth-claim (HARM), which I shall introduce just below.

⁴⁶ See my 1988, sec. III. As I have argued in my 1992, this difference between Socrates and Thrasymachus should have been clear from Socrates' reaction to Thrasymachus' complaint (337c2 with a8-c1) to Socrates that he is surely not going to say that justice is the advantageous, wherein Socrates replies that he would not be surprised if he did (337c9-10, 339a6-9, b3-4). And we do indeed find out that it *is* the advantageous – justice is what in itself makes us happy – it just isn't an advantage we gain by getting the better of others (*pleonexia*). For Socrates – and I think he is right about this – there is no such thing as getting more than others by harming them. I shall have more to say on this in future publications on the topic that might be described as there being no *pleonexia*, no science of gaining one's own good at the expense of the good of others – since, on the view equally of Socrates and Plato, there is *no such thing* as gaining one's good at the expense of the good of others.

⁴⁷ Two important clarifications should be made at this point. First, while Socrates holds that, as a matter of fact, one's own maximal happiness coincides with the good of those around one, this does not make the Socratic theory of human good 'just as much altruistic as egoistic.' For to be altruistic, it would have to be possible for there to be purely altruistic desires productive of action that are independent of the fundamental desire for one's own good. But there are no more such desires in the Socratic theory than there are categorical imperatives in the Socratic theory. (I am grateful to George Rudebusch for pressing me on this point, in a commentary he gave on a paper of mine given at Denver University in January 2000).

Second, since it is also important to my account of Socratic Ethics that it will never be beneficial to me to harm others, it might be thought that there is a little cheating going on here – as if the willingness to refrain from exploiting others is being used by me, without acknowledgment, as a *criterion* for a good ethical theory. The thought here would be that I am all unawares committing myself to the view that if it were possible to benefit by harming others, I would reject Socratic Ethics, and opt for morality over stepping on others. But this appeal to alternative possible worlds is entirely irrelevant to determinists like Socrates (as I construe him) and myself. For determinists do not believe in alternative possible worlds at all. All I am saying is that I reject any theory that has as a consequence the falsehood that it is beneficial to harm others. And a factual falsehood is *all* it is. Recognition of this point I claim to be a simple matter of human wisdom. (In any case, I do not myself believe, nor do I think Socrates would have believed, that morality as we understand it *does* correspond to the Socratic theory of goodness: see n. 50 below).

⁴⁸ It is true that, for most people, it is far easier to suppose that the claim in (11m) that it is morally wrong to harm others is true than it will be to suppose true the factual claim in (HARM) that it is in fact bad for you to harm others. The greater initial plausibility (11m) has may in turn tend to pressure one into accepting the exegetical claim in (11m) over that in (HARM). It is also true that, if what Socrates is doing in the *Crito* is claiming that we cannot be benefited by harming others, Plato does not, by our lights, do much by way of defending that claim. Prichard and Joseph both thought the view obviously false, at least in the unqualified form stated. I shall not be able to deal with that difficulty here.

⁴⁹ See my 1992a, esp. n. 22, with references back to my 1973a. See also Penner 1996, 1997 for the view that the climactic argument of the *Protagoras* (351b-360d – it should be thought of as a single argument) has the consequence that temperance (the 'measuring art') is identical with courage. As for justice, I admit that I am relying rather heavily on the account of justice in the *Protagoras*. But it can be argued that this account – wherein to speak of justice (as to speak of any virtue) is to refer to the *whole* of virtue – is confirmed by the ways Socrates speaks of justice in the *Lesser Hippias* and also in the function argument of Book I of the *Republic* (see n.16 above, as well as *Rep.* I.354e7-8 with 350c10-11, d4-5, 348c5-6: Justice *is* wisdom, and *is* the virtue of the soul). It is also confirmed by the paired arguments *Meno* 88b1-d3 and *Euthyd.* 279b4-8, 280a6-b8, 281c6-e5, each of which says that knowledge (*phronēsis* in the *Meno*, *sophia* in the *Euthydemus*) is the only

thing good in itself. I suggest this must be a wisdom which *is* justice, courage, temperance, and so forth. [This ‘thing good in itself,’ by the way, is the same ‘good in itself as a means to happiness’ that we have already encountered in n.14 above].

I shall not here address either the objection that I am wrong in supposing that Socrates, in the *Protagoras*, commits himself to the particular version of hedonism his argument employs there; or the following serious objection: that *Meno* 73a-77b and *Euthyphro* 12a-e seem to present to the reader, in the most unequivocal form, the assumption that justice is a part of virtue distinct from such other virtues as piety. On these two passages, I concede immediately that if we attend only to the immediate contexts of these passages, and omit consideration of what happens later in those dialogues, the objection will seem correct in both cases. It remains to consider whether what happens later in those dialogues makes a difference to how we see the objection. To this question I shall have to speak elsewhere.

⁵⁰ It is true that what I give here as the Socratic account of justice, thanks to the claim about harming others and one’s own best interests, has many of the same implications as the more usual accounts of justice, e.g., concerning stealing, breaking promises, harming others, and so forth. So it is possible to imagine that this Socratic account of justice comes to nothing different from the standard moral account. This is certainly not so. As just one example, take the harm that is done to some through punishment, even when those who have been convicted did intentionally the things for which they are being punished. Standard moral views have it that this harm is justified by the moral principle involved. This leads Plato to take the heroically implausible stand in the *Gorgias* that if punishment is to be justified at all – and of course most politicians will argue that it is – it will have to make the soul of the doer of injustice better (a possibility about which I express doubts in nn. 42 and 43 above). And in the *Rep.*, II.380ab, Plato declares that if it is not true that punishment benefits the soul of the offender, then no punishment is to be attributed to the gods. (The implication is that punishment may harm, in which case it is unjustified). But the idea that punishment benefits the punished is presumably not the view attributed to Socrates at *Apol.* 25c-26a, which, by implication, makes punishment never appropriate (again compare nn. 42 and 43 above) and so suggests Socrates could not see it as beneficial. (Compare also the same sort of view in Xenophon, *Mem.* I.ii.50, which, given the discussion of the distinction between ignorance and madness, is probably independent of the *Apology* passage).

⁵¹ Many modern readers of the *Republic* reject this (as I see it) straightforward parts-of-the-soul account of justice in which the just person is the person who acts with the knowledge had by his or her *rational part* as to what is best for *the person as a whole*. One basis on which this account is rejected involves the attempt to find in talk of ‘justice for its own sake’ a *moral* conception of justice. I have tried to sketch this view and my reasons for rejecting it above in n. 14.

Another basis on which the account is rejected is the claim that the Form of the Good directs us to suppose the soul that looks to the Form of the Good will desire an *impersonal*, non-relative good, not the soul’s *own* good. This idea (Irwin 1977, 1995, White 1979, Cooper 1977) is based largely upon a reading of the Forms descended from G.E.L. Owen, which, in my 1987, I have argued should be rejected. The Form of the Shoe that Simon the cobbler looks to is, I would agree, a non-relative Form, not, say, the Form of *Simon the cobbler’s* shoes. Against this, it seems to me that when Simon the cobbler looks to the Form of the Shoe in general, what he aims at is embodying that Form in *the shoes he will set out in his shop*. He is not trying to embody that Form in some shoe that is not in any shop, or in every shoe in every shop whatever, let alone (as noted in n. 16 above) in some general shoe. Similarly, if there were a Form of happiness, that too would be a non-relative Form – the Form of happiness in general – and not the Form of *my* happiness. That said, if I look to the general Form of happiness (or benefit), it will in general be because I want happiness to be realized *in my life*. That, surely, that is just what I do want. The *knowledge* I seek in order to embody it in my life is the knowledge of *happiness in general*. But my *desire* is for *my* happiness. My desire is that the real nature of happiness (happiness in general) be embodied in my life. So why should the fact that the Form of the Good is a Form of good in general make it the case that the person who looks to that Form does not look to it in order to secure his or her own good?

There is one other significant basis for the claim that we should reject the straightforward (factually egoistic) account of the just person as the person who acts so as to maximize his or her own good. This occurs in a much-contested – and crucial – passage at *Rep.* 519c8-521b11. The idea Plato is exploring in this passage – the passage is really no more than an unresolved parenthesis to the discussion of getting the guardians to take the upward path towards the Form of the Good (519c8-d2 with 521c1-3) – is the cute idea that a person is only fit to rule if it is not in the person’s best interest to rule (520d2-4). More or less the same idea was broached earlier in Book I, 345eff, esp. 347a-e; but it is there handled without embarrassment to egoism, in that such a ruler gets paid by means of the reward of not being ruled by someone worse. Here, on the other hand, Plato speaks of the guardians of the just city being *forced* to act *contrary* to their own maximal good or happiness by having to abandon the contemplation of the Forms in order to do rotations from time to time as rulers. What is more – note well: cutting against the force of the ‘compelled’ above – it is said that the guardians will find it *just* that they have to do this. Is this not, it is asked, the appearance of a *moral* good? Or, at any rate, is it not the appearance of a good – the good of one’s city – which is supposed to override the individual guardian’s own good?

Some interpreters think 519c-521b enough to *refute* the view that Plato holds the straightforward (factually egoistic) account of justice altogether. For reasons I shall shortly give, this seems to me to have the tail wagging the dog. If this passage *is* inconsistent with the straightforward account, then that can only show a strain in Plato’s thought about human goodness that is inconsistent with the egoistic strain in the straightforward account that he inherits from Socrates – and with the strain that says the just person is happiest. The former, non-egoistic strain would have been produced, perhaps, by an understandable longing in Plato the would-be *politikos* – the historical Socrates thought one shouldn’t waste one’s time in politics – a longing for a person whose aim would be the good of the city, even if that conflicted with the person’s own good. On the other hand, that the main line of the *Republic*’s account of justice does involve the just person seeking his or her own good seems to me undeniable (so that the best one can get from 519c-521b is the appearance of a certain unresolved tension in Plato’s view). What is more, the position in this passage is inconsistent not only with the relevant Book I passage, but also with the argument in the *Gorgias* (515e-521a) that the [political] science of making others better is the one science that needs *no further reward*.

This passage plainly deserves fuller discussion. Here I add only one further point: that 519c-521b is not about justice in the individual psyche – (JPSYCHE) in sec. IV above. It is about a detail in the founding (519c8, compare 420b6, 421a9-b1, c4), in imagination (472d9-e4, 501e4-5, 592a10-b5), of the ideal city, that founding being undertaken so that we can look to (JCITY) in discovering the truth about justice in the individual psyche. I have claimed, indeed, that Plato, in Book IV, has no interest in, and is committed to rejecting, (JCITIZEN), the view that citizens of just cities are *eo ipso* just individuals. We have seen that most citizens of the just city are not just individuals.

‘But,’ it will be replied, ‘the guardians in the just city *will* be just individuals by Plato’s parts of the soul account, and, what is more, Plato insists that these guardians put the good of the city above every other good, including their own.’

Not quite. What Plato insists on is that to make the ideal city work, he needs the guardians to put the good of the city above everything else (421b3-7, 420b4-8, c2-4, 421a7-8). Now here is the important point: That the guardians so act in an ideal city is the fact in the model that corresponds to the concern for *one’s own good* in the individual psyche. (One’s rational part looks to the good of one’s entire soul. The *guardian’s* concern for his or her *own* good that we are supposed to see refuted in the model is in fact an artifact of the model, with nothing in the individual corresponding to it. For the individual’s rational part is not another self).

Now Plato gives no sign of regarding this necessity that the guardians put the city’s good ahead of their personal good as a reason for rejecting the consequences of the model for the account of the structure of the just psyche, in which the rational part aims at the good of the entire individual. It would surely be excessive, therefore, to reject the main line account the *Republic* offers of justice in the psyche – that the rational part of

the just person aims at the good of the whole person – simply on the grounds that within the model, the guardian puts the good of the city ahead of every other good.

But does this artifact of the model still show a *tension* in the Platonic theory of the rational part's desire for the person's own good? That could certainly be argued. Much depends here upon how we read the many references to 'compelling' (mostly forms of *anankê*, necessity: 519c9, e4, 520a8, e2, 521b7, 421c1, 473d4-5, 499b5-c2, 500d4) the guardians to rotate contemplation with ruling, and how we interpret the many doubts Plato expresses about the 'possibility' of the ideal city (450c6-9, 458a2-b7, 466d6-8, 471c6-9, 472b1 ff, esp. 472c4-473b2, 499b2-c5, 500d4-e4). For example, is the point that in building the model, we *need* it to be the case that the guardians put the good of the entire city above everything else as we need the cobblers to be cobblers and not just feasters (420d5 - 421a2, compare 466b4-c3)? Are we, in building the model, just laying it down *as a condition of that construction*, that the guardians must do some ruling, and leaving aside the question how we get the guardians to fulfill that condition, as part of the larger question of the *possibility* of the ideal state? If so, then the artifact of the model will show nothing against the main line of the *Republic's* account of justice.

But I readily admit there are other possibilities. Later on in the dialogue, Plato seems to think more hopefully of the possibility of the ideal state if he can get something going by breeding the right natures (499b1-d6, 501d7-502c8, 540d1-e3, and 50e8-b5). At the same time the preposterous numerology of VIII.546a1 ff that underlies the possibility of a degeneration from an ideal city suggests Plato may even here be mocking the possibility of realizing an ideal city. As I have said, the passage, its context, and its bearing on the broader argument of the *Republic* all deserve much fuller discussion.

⁵² As I see it, (by contrast with Irwin and Bobonich: see n. 39 above), the only significant difference between Aristotle's 'three kinds of desire' version of the psychology of action and Plato's 'parts of the soul' version of the psychology of action is that actions brought about when appetitive or spirited desires trump rational desires are voluntary ('willing') according to Aristotle's doctrine, while according to Plato's they are involuntary ('unwilling').

⁵³ Accounts in the literature of the so-called 'Socratic Elenchus' all appeal to logical form of the sort I am saying we should not presuppose. These appeals to 'logical form' are akin to Rawls's liberal idea of saying what justice is by staying neutral on all metaphysical questions as to what the good is. This is, I think, yet another exemplification of the same sort of phenomenological approach that I find so unsatisfactory in meanings, logical forms, and the like. On *my* picture of Plato, he would have hated Rawls's idea. But I do not doubt that this suggestion of my picture of Plato will incur some opposition. For a little more on how I see the so-called 'Socratic elenchus,' see Penner and Rowe (forthcoming) and *op. cit.* in n.32 above. There is more on this whole topic in my *Plato and the Philosophers of Language*, and also in my 2004 De Vogel lecture.

⁵⁴ For convenience, this formulation carelessly quantifies over possible truths -- not something that, for other reasons, I am at all anxious to do. Put in more Fregean manner (cf. combinatory logic), wherein '(Ax)Fx,' i.e. 'A(F)' is a statement about the concept F, what is important is the issue of dependence: whether, as in (12a), what the correct account of the disagreement is depends upon the truth of the matter about which there is disagreement, or whether, as in (12b), it does not, and we can give an absolutely neutral account of the disagreement without any knowledge of the truth of the matter disagreed about – even knowledge as to which terms refer, and to what, in the sentences that formulate the disagreement. It is from this latter sort of platform that Sachs and Frege proceed.

In general, if there is to be meaningful disagreement about what is true of something (e.g., justice) then there must be a correct account of what the disagreement is about – so that justice will have to have the same reference for both of the disagreeing parties, whether or not they know what that reference really is, and even if they refer to it in different ways – possibly even in mistaken ways. (Such might have been disagreement about the sacred disease in pre-Hippocratic times. The disagreement was about epilepsy, whatever the superstitious belief of the disagreeing parties). There is no need that the disagreeing parties *know* even what the *way* is in

which their terms connect with the reference.

For every use of a term in a given context for picking out a reference, there will be a *way in which* it so refers in that context. But from this it does not follow – even for supposed canonical ways of referring (‘yellow’ to refer to yellow, ‘games’ to refer to games, and so forth) – that there is a *way of referring* such that *in all contexts whatever*, it is that way of referring that determines the reference. Language is not so regimented as we find it even in the later Wittgenstein’s comparatively more liberal approach, employing ever expanding language-games.

⁵⁵ It may be thought that I am here counseling that scholars throw away their LSJ and their Denniston. But the issue is not one of what reference books we use, but of what exactly it is that we *learn* from those reference books. Is it really the case that the altogether admirable Denniston divines what the words of a particular passage mean by intuiting meanings regardless of what he supposes the truth must be that the author in question is attempting to describe? If he intuits meaning-rules for various particles and particle-combinations, where does he intuit them from? By a visit to the author in the underworld? Or by a vision of the world beyond? Surely neither of these represents his method. As far as I can see, what Denniston does is to try to figure out in every passage what the author is saying really to be the case (or what the author is trying to make the case in the real world, or is asking about in the real world), and then inferring from *that* how we should take the particles or particle-combinations. He argues *from* (his best guess as to) what the truth is; he doesn’t first neutrally determine the meaning-rules and then ask whether or not what the author has said is true. Otherwise, it is far from clear to me just how Denniston decides that a particular usage is unique, or to be found in only two instances.

⁵⁶ One might argue: I know the *meaning* of ‘justice’; so I can understand the question whether justice is more profitable without knowing what the answer is to the question ‘What is justice?’ But this is just to fall back into the Fregean approach. If there are any meanings at all, then perhaps by knowing the meaning of ‘justice,’ I can come to know the *meaning* of a question about justice. My point has been that this does not get you knowledge of *what the question is*.

⁵⁷ According to Geach, the Socratic Fallacy is thinking both that

A If you know that you are correctly predicating a given term “T” you must “know what it is to be a T” in the sense of being able to give a general criterion for a thing’s being T,

and also that

B It is no use to try to arrive at the meaning of “T” by giving examples of things that are T

To see how this is supposed to be a fallacy, we have to realize that (A) knowing you are correctly predicating the term ‘T’ is knowing the meaning of ‘T,’ and giving a general criterion for a thing’s being ‘T’ is giving *as* the meaning a quality common to all instances of T. (This is of course the common quality notion of meaning that is the *bête noire* of Wittgenstein’s *Philosophical Investigations*). It is Geach’s view that we know what we mean by all sorts of words – any words we employ in ordinary use – without having any common quality in hand. The contrast between our knowing the meanings of ordinary words and not yet knowing the quality common to all uses of the word is very close to the view in Prichard 1928: see 1968, 207, 215-7). So understood, what (A) says is that we don’t know the meanings of the words we are using in the questions we are asking until we have a common quality; but (B) says we cannot arrive at the common quality by the only way Geach thinks is open to us, namely, by way of examples (since we don’t know what are examples till we know the meanings of the words we are using in our questions). The error is of course supposed to lie in the common quality theory of meanings. We know the meanings of our words without having as criteria common qualities. As Ryle would put it, we *know how* to use our words, without *knowing that* we use them in accordance with such and such a general criterion. (It would be interesting to know whether in the

conversations between Ryle and Wittgenstein *ca.* 1930, they ever discussed Prichard 1928 – which Ryle will certainly have known of).

Geach is quite wrong to suppose that Socrates is committed to either (A) or (B), since – this paper urges – Socrates would deny that meaning determines reference, and so would deny the existence of meanings altogether. Geach thinks – in accordance with the Fregean theory of questions – that we have to start with meanings. It is a consequence of the main point of the present article that there is *no such thing* as the fallacy Geach claims to diagnose. Indeed, turning away from the *meaning* of ‘justice’ to the real nature of justice, Socrates is quite right that we cannot know whether justice will make us happy or not without knowing what justice is.

⁵⁸ Just as I was finishing up the present version of the paper, I came across something I had read much earlier – about Gadamer on propositional logic, statements, question and answer, and context. As I understand it, the suggestion is that we need to attend not to the philosopher’s statements but to the questions he is trying to answer, where questions are understood in terms of context. I now see that this suggestion may be very close to the kind of thing I am suggesting here. And this led me to wonder why I had not noticed this similarity earlier. The probable reason is that I did not when I read this material earlier see how this suggestion would help if one had no alternative to Frege’s theory of questions. Until one refutes this theory of questions, it does not help to change the subject from statements to questions. I don’t know how Gadamer’s account would stand on my present arguments against Frege’s theory of questions, or on the matter of whether Thrasymachus intended to refer to the real nature of justice, even if *what that is* is quite different either from how Thrasymachus (or for that matter anyone else) conceives it. See P. Christopher Smith’s preface to Gadamer 1986 [1978], xi, xii.

⁵⁹ The present paper involves, except in nn. 1 and 28, insubstantial revisions to the paper of summer 2001 referred to by Santas 2003. Earlier versions of this paper were read to a graduate seminar in Madison in the fall of 1999; at the Texas Association for Classical Philosophy in Forth Worth, March 2000; at the 13th International conference of the Olympic Center for Philosophy and Culture, ‘The Political Theories of Plato and Aristotle,’ at Olympia, Greece, August 2000 (supported by a travel grant from the University of Wisconsin Graduate School); and at a graduate seminar divided between two campuses of the University of California (San Diego and Irvine) in the spring of 2001. Most of the material was hatched much earlier. For example, the main problem I faced (secs. III, VIII) after dispatching Sachs’s argument was worked out during a long automobile trip with Jerry Santas in February 1997. Earlier still, material earlier in the paper was discussed with Christopher Rowe in a long afternoon at O’Hare airport. The penultimate draft of the 2001 version of the paper profited greatly from discussions with Dennis Stampe on causal theories and on meanings, and with James P. Butler on the fundamental question of the *Republic*. It also received a searching critique from Ruth Saunders, and a long conversation with Ruth also helped me greatly in getting this fairly complicated exposition under somewhat better control. I am very grateful for all the help I received from all these sources. I am also grateful to Christopher Rowe and Gretchen Reydam-Schils for their help at a recent critical stage in the paper’s history.