

HOW PLATO REASONED ABOUT JUSTICE IN HIS *POLITEIA*

Gerasimos Santas

University of California, Irvine

In his important book, *Theories of Justice*, Brian Barry gives credit to Plato for starting political philosophy, by raising the question, what is justice? I think this is correct and the credit is well deserved. Barry goes on to reject what he takes to be Plato's own answer to this question, the theory of justice Plato puts in the mouth of Socrates. But what he and others do not appreciate is that Plato presents us with three different major theories of justice, the theory of Thrasymachus (T), the theory he puts in the mouth of Glaucon (G), and the theory his Socrates (S) elaborately constructs. Even more important, and even less appreciated, Plato teaches us how to reason about what justice is: his T and G and S not only give different answers to the same question, but look for different things and search in different ways in order to reach the different answers. Plato teaches us how to think about justice. This is something we can appreciate and learn from, even if we disagree totally with the answer Plato has S construct. This is the subject of this paper.¹

In the *Republic* we have at least three different major theories of what justice is, arrived at by three apparently different methods of reasoning about justice. The significance of methods of reasoning about justice becomes clear: different methods might give us different results, as they actually do in the *Republic*, and then perhaps we can better understand, even gain some control over the differences; on the other hand, if different methods gave us the same result, our confidence in that result would reasonably be greater.

If we need confirmation for the significance of methods of reasoning about justice we can find some in John Rawls, *A Theory of Justice* (1971. Ch. III, section 30). He points out that in the Original Position the parties would choose the principle of average utility over the classical principle of total utility; he notes that the supporters of the latter knew this, and then in remarkable discussion Rawls finds a different method and foundation for the principle of total utility: the device of the ideal observer and the application of the principle of maximizing an individual's overall good to society.

Of course Plato did not have the benefit of the history of ethics and political philosophy, which Rawls knew well. Plato was a pioneer. Moreover, he used an indirect and informal dialogue style, and whether from ignorance, uncertainty, or choice, he wrote in an Aartfull chiaroscuro, an interplay of light and shadow, voice and silence. We must be clear about what he explicitly says and try to light the shadows and voice the silences by hopefully educated guesses.

The Empirical Method of Thrasymachus

When asked about the burdens of old age and the uses of wealth, the wealthy old Cephalus naturally thinks of wealth as a means of restitution and clearing his conscience, if he has injured anyone in this life by lying or breaking a promise. Socrates puts these two rules in Cephalus' mouth as an account of justice and brings up counter examples to dispute it. Though the conversation with the old man foreshadows many themes of the *Republic*, there is no method here of arriving at this account of justice, other than Cephalus' circumstances and experience.

Nor is there a method of reasoning about justice in Polemarchus' appeal to the poet Simonides, for the view that justice is rendering to each man his due, harm being due to enemies and benefit to friends. It is simply an appeal to authority. The poets invoke the muses for inspiration, and Polemarchus appeals to the poets for justice.

By contrast, Plato explicitly portrays T as employing a certain method for arriving at his account of justice as the advantage of the stronger:

“Don't you know then, said he, that some cities are governed by tyrants, in others democracy rules, in others aristocracy? Assuredly. And is it not this that is strong and has the mastery in each B the ruling party? Certainly. And each form of government enacts the laws with a view to its own advantage, a democracy democratic laws and tyranny autocratic and the others likewise, and by so legislating they proclaim that the just for their subjects is that which is for their B the rulers' B advantage and the man who deviates from this law they chastise as a lawbreaker and unjust. This then, my good sir, is what I understand as the identical principle of justice that obtains in all states, the advantage of the established government. This holds power, so that if one reasons correctly, it works out that the just is the same everywhere, the advantage of the stronger.” (R. 338e-339a)

T begins with an elucidation of the term “stronger” in his initial definition (“justice is the interest of the stronger”): (1) the ruling party in each form of government, whether a tyranny, a democracy, or an aristocracy, is the stronger. Next, he asserts a big empirical generalization: (2) that in each form of government the ruling party enacts laws to its own advantage; the laws may be different in different kinds of government but they all have this feature in common. He then brings in an assumption: (3) that each form of government proclaims that justice consists in observing the laws it has enacted and injustice in breaking these laws; justice in a society is found in its laws and nowhere else. And finally, T concludes from these three premises that (4) justice is everywhere the same, the advantage of the stronger.

This is a method for determining what justice is: on the assumption (3), that in each society the laws determine what justice in that society is, the method consists in an empirical investigation of the aims of the laws and the motives of legislators in each society, and then generalizing from the results to what is common (if anything) to the justice of all societies. On the same assumption, anyone can perform this investigation of the constitution and laws of, say, Sweden and find out what justice is in Sweden, or in India, or any other society or form of government. It is an investigation which now days would be done by an empirical political scientist doing comparative government.

T did not actually do the empirical research, he only announced its alleged results. But Aristotle apparently did such an empirical investigation of many different constitutions (some 154 constitutions according to some sources) and found that in some cases (we don't know how many or percentage of the total), T's result does indeed obtain (*Politics*, Bk. III, Chs. 4,5). But he disagreed with T's conclusion that all constitutions have this common feature of aiming at the advantage of the stronger or ruling party; and he certainly argued that none of these constitutions are just. Plato himself foreshadows Aristotle's conclusion in the *Laws* (Bk. IV, 714-15; Bk. III, 697).

S's first argument against T (*R.* 339) confirms the crucial importance of the assumption that makes the empirical investigation possible: that justice in each society is determined by the laws of that society. R.E. Allen has called this assumption "legal positivism" and claims that the American Justice O.W. Holmes adopted it from T (Allen, 2006, ix). This assumption is based on the plausible idea that justice can be found where there is law; but it goes beyond this idea by claiming that the only justice that exists is to be found in the positive laws of societies; so if we want to find out what justice is we investigate these laws.

S points out that rulers or legislators might sometimes make a mistake in supposing that a particular law they enact would be to their own advantage, and in such a case acting in accordance with such a law would be both just (by the assumption in question) and also unjust (by T's definition of justice as the advantage of the ruler). This is correct, and has a more general validity: so long as laws are thought of as means to some end, such empirical mistakes are possible, whether the end is the advantage of the ruler or the advantage of every citizen or the common interest. Thus the assumption would be incorrect even if we defined justice as what is to the common advantage or the advantage of every citizen; even in that case it would be incorrect to claim that the laws of a given society completely determine what is just in that society. The possibility of such mistakes shows that laws can be unjust. This argument clearly opens up some logical space between positive laws and justice.

Socrates' second argument against T's account of justice (*R.* 341-3) confirms that the method is an empirical investigation and the main premise (2) an empirical generalization. Socrates argues that if ruling is an art or science, we can compare it to other human arts (such as medicine and navigation, for example) and see whether in such cases the aim is the advantage of the practitioner (the stronger) or the interest of the subjects (the weaker) over whom the art is practiced (healing of the patient, and safety

of the passengers at sea). He concludes that just as in the case of medicine and navigation the aim of the art is not the advantage of the practitioners but that of the subjects over whom the arts are practiced, so in the case of ruling, contrary to T's generalization. S's argument has the form of an analogy, which is the right kind of argument for testing T's empirical generalization, short of carrying out an actual investigation of different kinds of constitutions, as Aristotle did.

We see then that Plato has T use a method of investigating the nature of justice (what justice is), he has S dispute the assumption which makes that method possible (that the laws of a society completely determine what justice is in that society), and he even has S dispute the empirical result of the investigation even if the assumption is granted (by T's restriction of a ruler to one who makes no mistakes; this restriction is not really helpful to T, because there are not rulers who never make mistakes, and so T's justice, so restricted, has not application. *R.* 340).

Seeing clearly the method by which T reached his result enables us to assess that result critically, by examining its basis, the premises and the reasoning by which the conclusion is reached. We can even tell how one premise, the empirical generalization, might be confirmed or shown to be false: by examining the constitution and the legislative practices of each society or form of government.

The dispute about the other premise (3) is fruitful too: by treating laws as means to ends Socrates opens up some logical space between law and justice: we can't suppose that laws are always just, no matter what ends laws are supposed to serve, because human beings can make mistakes about means to a given end. Our empirical political scientists can still do interesting empirical investigations of the justice embedded in each form of government, but it has to be on the weaker assumption, that what a society *thinks* justice is can be found in its constitution and laws, an assumption that leaves it open that a whole society might be mistaken and its justice is really injustice. T's super conclusion, that justice is the interest of the ruling party, will never follow; but at most, that what is thought to be justice turns out everywhere to be the interest of the stronger party, a disturbing enough conclusion. By opening up some logical space between justice and law Plato was able to dispute the justice of some laws even if they are universally present in all societies, as, for example, the laws determining the inequality of women, which were universally present in Plato's world.

The Contractarian Method of Glaucon

The speeches of G and Adeimantus enrich the *Republic* remarkably. Social contract theories of justice had hardly been born when Plato wrote the book. In one short paragraph Aristotle attributes one to the sophist Lycophon (*Politics*, Bk. III, 9, 1280,) and its significance seems to escape him (For a recent discussion see Keyt and Miller, 2004). It took some twenty centuries for such theories to make a second significant appearance, but then they had a most fruitful run from Hobbes to Rawls.

A contractarian theory of justice claims that justice is the object and the product of a voluntary and presumably rational choice or agreement among human beings;

and it usually supposes that before there is such an agreement there is no such thing as justice. Contractarian theories must have at least two main parts: the conditions under which the agreement or choice is made B usually called the circumstances of justice B and the content of the agreement (the alternative chosen among their options) that is reached voluntarily and rationally in the circumstances described. The conditions under which the agreement is made, the options before the parties, and the reasoning used to reach some agreement may be called the contractarian method.

G's account of justice is contractarian and it is usually so taken. He tells us explicitly that the first thing he is going to do is to give "...the nature and origin of justice", and he starts with the origin:

"... by nature, they say, to do injury is good for one, to suffer injury bad, but the bad of being injured exceeds the good obtained from injuring others. So that when men injure and are injured by one another and had a taste of both, those who lack the power to avoid one and do the other determine that it is in their interest to agree with one another neither to injure nor to be injured; and this is the beginning of legislation and covenants among men, and they name what the law commands the legal and the just, and that is the origin and the nature of justice. It is a compromise between the best, which is to injure with impunity, and the worst, which is to be injured without the power to retaliate." (R. 359, Shorey transl.)

Here we find the two main parts of a contractarian theory: a state of affairs in which men presumably had no justice, and were injuring and harming each other and suffering the consequences; and the agreement they reached: each agrees not to injure others in return for a similar agreement by others not to injure him. The first part gives us the circumstances in which justice originated or was created by agreement, and the second part gives us the content of this fundamental agreement and the reasoning that led up to it. Though we can see here these two parts of a contractarian theory roughly and in outline, many pertinent questions are not answered in our texts, and we can only try to make educated guesses about answers Plato would have given.

In the state before justice, the state of nature as the moderns call it, what was the environment like? Since it seems that men found it necessary to injure and harm each other, presumably as a means to getting the things they wanted (and not for its own sake), we can infer that these things were not in abundance B that it was a state of moderate scarcity in the things men usually need and want. Perhaps it was a state represented by a zero-sum game: if we sum up all the transactions among men B voluntary or not B the gains and losses sum up to zero, and normally one man's gain is another man's loss. Some have suggested that it might even be a negative sum game (since G says that "the bad of being injured exceeds the good obtained by injuring others." R. 359).

And what were the human beings like in this state? Apparently they were self-seeking, whether completely or predominantly so: each seeks the things which presumably he thinks he needs or wants or regards as good for him; including apparently injuring or harming others, at least as a means to getting what he wants, and appar-

ently in the expectation he will not be retaliated upon or that he can successfully repel retaliation or that he can come out ahead even with retaliation..

This for their circumstances and their motivation. But what of their capacities? Apparently, they were also minimally rational: at least able to learn from their experience in the state of nature, and able to figure out effective means to their own ends and their overall good. Thus, moderate scarcity in the things they want, their self-seeking nature, and minimal rationality, seem the minimum assumptions necessary to account for the conflicts among them, and their agreement with each other.

And what do they agree on? What is the content of this first and fundamental agreement? G says, “neither to injure nor to suffer injury.” Apparently this means at least that each agrees not to injure others (or to give up his freedom to do that) provided that others agree not to injure him (or others give up their freedom to injure him). Once they set up laws in accord with this fundamental original agreement, just conduct will be determined by such laws; and sanctions for disobeying them presumably can be expected to give to each the security of not being injured by others in return for the freedom s/he has given up.

G also tells us that the justice that comes about by this agreement is regarded as a compromise between “the best, to injure with impunity, and the worst, to be injured without the power to retaliate.” Apparently this means that the parties to the agreement think that the best state of affairs for them is to have the freedom to do whatever they please including injuring others, together with the power to do so with impunity; and the worst state of affairs is to be injured and not to have the power to retaliate.

This interpretation of G’s account of the origin and nature of justice, though minimal, is still full of inferences from our texts. We inferred moderate scarcity, we inferred at least predominant psychological egoism, we suggested a zero sum game. We inferred rationality about one’s overall good. We amplified the content of the agreement. Further, we guessed that the principle of equality later mentioned by G (*R.* 359c) refers to the content of the agreement: giving up the same freedom equally in exchange for the same security equally. We softened the extreme emphasis on injuring or harming others B we described it as a necessary means to one’s perceived good (not as something pursued for itself). And we amplified the best and worst states that the agreement forbids and avoids.

Even so, several important questions remain unanswered. What alternatives did men have to choose from? Only the state of nature and what they ended up agreeing on? Why so? The contractarian method is open on what the alternatives are in front of the choosing parties. Even within the limited philosophical space of the *Republic*, we can imagine different principles of justice, and different constitutions and forms of government the parties could have considered: T’s principle of justice, Plato’s own principle of social justice, as well as the one they actually agreed on. Or, more concretely, they could have considered a choice among democratic, oligarchic, timocratic, even tyrannical constitutions. And if they had all these alternatives, they might have to use more complex reasoning to make a choice, and perhaps take into account odds

as well as outcomes.

Another important question left unanswered: was the agreement unanimous? Our texts do not say, though conceivably they imply that it was. The problem is that G says: “those who lack the power to do the one [injure or harm others] and avoid the other [being injured by them to even a greater extent]... agree with one another...” This suggests that not everyone was in that situation of weakness: perhaps there were some who had the power to injure or harm others and get away with it. Did they agree? Why would they give up the freedom they enjoy for a security they already had? We may have to suppose a two stage agreement: first, those individually weak agree among themselves for the reasons given; once they band together, being perhaps a considerable majority, they are collectively stronger than the few individually strong men, and then the latter find themselves now in the weak position and also come to agree for similar reasons.

This interpretation perhaps accounts well for some of the things that Plato has G say in the rest of the speech: G claims that anyone in the created city-state would behave unjustly if given the magic ring of Gyges, which makes one invisible and able to avoid punishment. But he also implies that only few in the state would have the societal equivalents of the ring, the ability to act unjustly secretly and to cultivate the deception of a just reputation. In turn, this seems to serve well Plato’s implied criticism of G’s contractarian justice: in the city state which would result from Glauconian agreements, the best and brightest individuals would be dissatisfied and tempted to try to escape the bounds of the agreed upon justice B they would represent the strong in the state of nature and the least benefited by the original agreement. G’s just city might be highly unstable. A justice which tempted the best and the brightest under it to act unjustly leaves something to be desired.

Finally, there is the question of what are the goods or the interests for the sake of which men fought in the state of nature, and for which the best and the brightest might be tempted to act unjustly in the resulting society. In G’s thought experiment, when just and unjust men are both given magic rings, they equally go after power, property, wealth, and pleasure; they are portrayed as happy because they get these things, and more generally because they can do and have whatever they desire or please (they conduct themselves as equal to a god! *R.* 360c). So presumably it is for the sake of these things that they fight in the state of nature, and for the sake of these B or at least to avoid their opposite evils B that they compromise and agree, given that they lack the powers of gods.

As we learn later, Plato’s S disagrees that these goods make people happy, and proposes other goods that out rank them, such as knowledge and health and the pleasures of good activities. Rival theories of what is good for human beings, as well as rival theories of justice, are in play in the *Republic* – one of the things that makes the book a fundamental work.

The Functional Method of Socrates

Plato does not have S criticize directly G's theory of justice; he does not unleash his star critic on his brothers, as he had unleashed him again T. Instead, he has S start a new inquiry into what justice is. But he does not tell us explicitly by what method he is trying to discover another answer yet to the question what justice is.

If we track his actual investigation from that point (R. 367) all the way to his two definitions of a just city-state and a just person (R. 432, 442), we can plainly see that there is no trace of T's empirical method, nor of G's contractarian method. Nor is there a trace of a Socratic method we might extract from the earlier dialogues of definition: collecting some clear examples of just actions or just persons, generalizing from them, and testing the generalization by further examples and by consistency with other firmly held beliefs. (The contemporary analogue of this Socratic method so conceived is, I believe, Rawls' method of "considered judgments in reflective equilibrium", Rawls, 1971, ch. I, section 9).

Yet we can hardly suppose that Plato reaches his specific results randomly or by sheer good fortune. And since he does display considerable sophistication about philosophical methods, not only in the views of T and G but earlier too in the Socratic dialogues, we must at least make a search for his method. Can we discover a method from what he actually does between his starting point and his definitions?

Well, Plato has Socrates begin by dividing the question into two: What is a just city, and what is a just person? Apparently Plato takes these two cases, the justice of a city and the justice of a person, to be the primary or central applications of justice. Of course he knows that justice has other important applications, for example, to laws and actions; but apparently thinks that if he can discover what justice in the city and in the individual is, the rest of the cases can be dealt with more easily.

S starts with just city (368c), and supposes that if they imagine a city coming into being they would see justice or injustice coming into being too. He suggests that the city comes into being because humans, being individually not self-sufficient, come together to render services to, and trade with, each other, "because each supposes this to be better for himself... let us create a city from the beginning in our speech... its real creator will be our needs." (R. 369c). In rapid order S (1) proceeds to list the economic needs for food, shelter and clothing, (2) proposes a division of labor for producing these goods, (3) supposes that human beings are born with different abilities for different kinds of occupations, and concludes that "... more things are produced, and better and more easily when one man performs one task according to his nature, at the right moment, and at leisure from other occupations." (R. 370c. Note that S does not claim that each person can practice only one kind of occupation, but only that one cannot practice *well* more than one, given that different inborn abilities and education and leisure and timing are necessary for doing such things well. As inborn ability, education, and time are required, "it is impossible that one can practice well many arts." (374, 394c). In this last passage it is clear also that it is not minute division of labor that is at issue (as in Adam Smith's, *The Wealth of Nations*, Chs. 1-3), but division into crafts,

the arts and sciences.

In successive passages S expands the needs of citizens of his imaginary city, beyond provisioning themselves, to defending and governing themselves (373d-74e); tries to figure out what inborn abilities are best suited for which of these several social tasks (374e-76e); suggests a system for educating these several abilities (376e-415db); proposes institutions of property and family for the citizens whose inborn abilities and education suit them to defend and govern (415-27); and ends up by claiming that the city they have imagined, if they made no mistakes, is “completely good.” (427e).

He immediately infers from the city’s complete goodness, that it is “wise, brave, temperate, and just.” Finally, he proceeds to give an account of each of these four virtues in the order just given, trying to catch the particular good that each virtue contributes to the city. He ends up with the claim that the justice of the city is to be found in the principle of organization they started the city with: “For what we laid down in the beginning as a universal requirement when we were founding our city, this, I think, or some form of this, is justice... that each man must perform the one social service in the state for which his nature was best adapted.” (R. 433a).

Now why should Plato suppose that by following this procedure he would discover what justice in the city is? Indeed, why did he suppose that by following it he would discover what a completely good city is? And why did he think that he could infer from its complete goodness that the city had these virtues?

I think we can find convincing answers to these questions in a procedure suggested by a theory Plato has S expound and use at the end of the first book (R. 352eff).

The theory first gives an account of two kinds of functions: “... the work (*ergon*) of a horse or anything else [is] that which one can do only with it or best with it...”; and again, “... when I asked whether that is not the work (*ergon*) of a thing which it only or it better than anything else can perform.” (353a7). I call the first kind of functions exclusive and the second optimal. S gives seeing as the exclusive function of eyes (since only with the eyes can we see), and hearing as the exclusive function of the ears (since we can see only with the ears); and he gives pruning as an optimal function of a pruning knife, since, he says, “we can use a dirk to trim vine branches and a knife and many other instruments... but nothing so well as a pruning knife fashioned for that purpose.” (353a).

Second, the theory proposes that there is (an) appropriate virtue(s) for each thing that has a function, and characterizes the appropriate virtue(s) of a thing with a function as that by which it performs its function well, and its vice(s) that by which it performs it poorly. (353bc)

Next, S claims that this theory applies to “all other things” (presumably to all things with functions), and immediately applies the theory to prove to T that a just man is happy and an unjust one unhappy. (353d-354a). This immediate application proceeds by supposing that the soul has the exclusive functions of managing, ruling, and deliberating, and that justice is the virtue of the soul and injustice its vice; from the theory and these premises Socrates then concludes that the just soul will do these functions

well, the unjust one badly, and that the soul that does these things well will live well, the unjust will live badly; and further that the soul that lives well is happy and the one that lives badly unhappy.

During this application S adds to the theory: he says that a soul that has the virtues of a soul will be a good soul (and one that does not have them a bad soul), and that a good soul will perform its functions well, a bad one poorly. (353e). Presumably, generalized versions of these two premises can be added to the theory: a thing of a certain kind (with functions) that has the virtues appropriate to that kind will be good of its kind; and a thing good of its kind will perform the functions of that kind well. Arguably, in this application he also needs the assumption that functioning well is good for the thing that functions well, perhaps its chief good; seeing well, for example, is the good of the eyes, and anything that contributes to seeing well is good for the eyes.

I think we can see from the examples that S gives that the theory generalizes from practices, including evaluative practices, in medicine and the productive arts. The medicine of Plato's day had determined that the human body has a natural division of parts (especially organs, but also fluids, tissues, and bones – see the long discussion of the human body in the *Timaeus*, 72-79); and a natural division of labor matched (or assigned by the divine craftsman, in Plato's view) to those parts; eyes and ears are natural parts of the human body and each has a unique task matched to it; we can understand the human eye by understanding what that function is and by finding out what qualities, such as structure and composition, enable it to perform that function well. Further, we evaluate the human eye by how well it performs that function, and we can think of the qualities that enable it to perform that function well as its virtues. It is then a truism that an eye that performs its function as well as possible is a completely good eye and that a completely good eye has all the virtues appropriate to eyes (i.e. the virtues relative to that function). The theory for human artifacts is the same, except that the account of function is different: any artifact can be used for many purposes, but usually it is the best instrument (better than others) for the purpose or use it was designed for (rather than other uses); and it is good of its kind and functions well when it has the virtues appropriate to its function. (For a critical discussion of the functional theory, see the author's *Goodness and Justice*, pp. 66-75. For discussions of function in biology see Hull and Ruse, 1998).

This theory suggests a procedure for discovering the virtues of objects of a given kind, on the assumption that such objects have functions: (1) find out what the functions of such objects are; (2) determine (by observation, experiment, or even thought experiment) cases where objects of such a kind perform their functions well and cases where they perform them poorly, and (3) finally find out their qualities which enable them to perform such functions well (and in the absence of which they perform poorly), and these are their virtues. (For procedures consisting of several ordered steps, which Plato calls a *methodos* (method), see *Phaedrus* 270b-71c, *Sophist*, 243d-244, and *Republic* 532ff).

And we can now see that this procedure accounts well for S's major moves in try-

ing to determine what justice in a city is. On the assumption that justice can be found in cities (as well as in individuals), (1) he tries to discover what the functions of such city-states are B to provision, to defend, and to govern themselves. Next (2) he tries to organize an imaginary city-state so as to perform these functions as well as S and his interlocutors can think of; by division of labor matched to a division of citizens by natural ability for these functions and suitable education. Finally, (3) S appropriately claims that if they made no mistakes in the way they divided, structured and educated their imaginary city (so as to perform its functions as well as they can think of), their city is indeed completely good; if it is completely good it has all the virtues appropriate to the city; and since justice is a virtue of city-states, their city will have justice, and now they can try to locate it among the qualities which enable their city to perform its functions well and be a completely good city.

This procedure is by no means complete. If, for example, a thing has many functions and many virtues, it does not tell us how to differentiate among its several virtues. Plato claims that both city and soul have several functions and several virtues; but which virtue accounts for the well performance of which function, so that the virtue can be characterized accordingly? We can see his hesitation in trying to define the virtues of the city: he lists them in a certain order, leaves justice for last, as what is left over after wisdom, courage, and temperance have been defined.

He has an easy time with wisdom and courage, since it is clear to what functions each contributes (making good city decisions, and defense the city well; 427-30); but struggles with temperance, which a virtue of the whole rather than any one part of the city (he goes first to temperance as it applies to individuals, and finally defines it as a city virtue by combining control of the best part over the worst and lack of conflict among the parts, 430d-32-e); and when it comes to justice, which also is a virtue of the whole city, he feels he has to find arguments that what he has defined is the virtue of justice: assigning city functions to parts of the city optimally, that is, on the basis of what city function each city part can do best, may help the city perform all its functions better, but why is this justice? (432b-34c; and Keyt, 2006) Despite these problems, in his accounts of the city's virtues we can see that in each one of them S tries to catch some quality of his imagined city which would account for performing some one or all its functions well. Beginning with the less controversial cases, knowledge of what is good for the city as a whole is said to account for the city being well governed; its courage for being well defended; while temperance would account for the good of harmonious relations among the parts of the city, especially on the question of who should govern the city; and justice accounts for the good of all three functions being performed better than they would be under any other political and economic structures (i.e. a timocratic, oligarchic, democratic, or tyrannical constitution).

Now there are some who think that the first book is a Socratic dialogue whose theories cannot be relied on to interpret the rest of the work. They think this applies to the theory of function and virtue of the first book, and cite Socrates remark at the end as specific evidence that the theory of function (ergon) is abandoned there. (Perhaps

Vlastos, 1991, Burnyeat, 2002).

I certainly disagree with this general view; I think the first book foreshadows many important themes in the rest of the work and is beautifully integrated with it; but this is beyond the scope of this article. However, aside from this general dispute, I think we have convincing evidence that the theory of function and virtue of book I is used to give an account of justice in the rest of the work.

First, S's remark, at the end of Bk. I, is not good evidence that he abandons the functional theory. He says that, like a glutton, he rushed to discuss whether justice is a virtue, and whether it is better than injustice, before finding out the first object of the inquiry, what justice is. "So that for me," he continues, "the present outcome of the discussion is that I know nothing; for if I don't know what justice is, I can hardly know whether it is a virtue or not, and whether its possessor is or is not happy." (354bc). I think this remark is well accounted for if we suppose S to be saying that his arguments against T on the benefits of justice are unsuccessful because S had no account of what justice is: he argued that justice is a virtue without saying what justice is; and again he argued at the end of the first book, using the functional theory, that justice makes one happy, and injustice unhappy, without saying what justice is. We cannot obviously suppose that S was arguing that justice *according to* T (i.e. the advantage of the rulers) is a virtue and that it makes its possessor happy, since S does not think that this is what justice is. So, S is correctly complaining that in effect he has been arguing that justice, *whatever it is*, is a virtue and makes its possessor happy. But how on earth could anyone possibly show or know *that*? In any case, his remark is no reason to throw out *all* the premises in those two arguments but only reason to try again, this time taking into account the priority rule of giving an account of what justice is before proceeding to show that it makes one happy. And this is exactly what S does in the rest of the work.

More important, we have good textual evidence for thinking that Socrates is using the functional theory in the rest of the work.

S uses the term *ergon* (function, work) significantly several times when he discusses the origin of the city and states the principle of social justice for the first time. Thus when he first proposes division of labor he asks: "Shall each of these [the citizens] contribute his own work (to *outou ergon*) for the common use of all?" (369e). Next, he uses the term when he introduces the natural lottery assumption: "... our several natures are not alike but different... one man is naturally fitted for one task, another for another (*allos ep allou ergon praxein*)." (370b). Again, he uses the term to point out that the various occupations require leisure from other things so that the work can be done at the right time (*ergou kairon*, 370b). Again, when trading is found necessary for the city, Socrates says that those weakest in body and useless for any other function (*ergon*) will be traders. Looking back at the account of *ergon* (function) Socrates gave in the first book, we can see clearly that the concept of optimal function is being used here: citizens are to be assigned that *ergon* for which they are best suited by nature and (added later) appropriate education.

Next, when S introduces the need for defending the city, he extends the scope of

the principle of social justice (whose point is restated at 374b in terms of *ergon*—“to the end that the cobbler’s *ergon* is well done”) to the function of defending the city: “Then, I said, to the same degree that the task of the guardians (*ergon*) is the greatest of all, it would require more leisure than any other business and the greatest science and education. Does it not also require a nature adapted to that very pursuit?” (374e) It is clear in this passage again that the concept of optimal functions is being used.

In Bk. III, 406c, *ergon* is used in a version of the principle of social justice attributed to Aesculapius: “... he knew that in all well governed peoples there is a work (*ergon*) assigned to each man in the city which he must perform....”

In book IV, we find *ergon* explicitly linked to the typical formula which Plato’s uses for city justice and soul justice in that book (*to eautou ekaston pratten*), each person does his own. In replying to Adeimantus, S says: “... but these helpers and guardians are to be constrained and persuaded to do what will make them the best craftsmen in their own work (*tou outon ergon*), and similarly with all the rest [of the citizens].” (421c) And a bit later, in the discussion of the unity of the city and of each citizen, S says: “... the other citizens too must be send to the task for which their natures were fitted, one man to one work (*ergon*), in order that each of them fulfilling their own may be not many men, but one, and so the entire city may come to be not a multiplicity but a unity.”

Erga (functions) is also used in the beginning of the final argument S gives to show “that each doing one’s own” is what justice is, and two kinds of injustices violations of it: “A carpenter undertaking to do the work (*ta erga*) of a cobbler or a cobbler of a carpenter... or even the attempt but the same man to do both... would not greatly injure a city?” (434a-c). Thus the two kinds of injustice, practicing an art one is not best suited by nature and education and practicing many arts, are described here in terms of *ergon*; once again clearly the notion of optimal function. These two cases of injustice account well for the four types of unjust city and unjust individual in Bk. viii.

Now one might grant that Plato uses the three step functional procedure to discover justice in the city, but deny that the functional theory plays any role in the account of the virtues of the soul. And it is true that for discovering justice in the individual soul Plato does not explicitly use this functional three step procedure, but rather a shortcut made possible by his assumption that a just soul is isomorphic to a just city, a shortcut that enables him to deduce an account of the just soul from his account of the just city and an independent argument for his division of the soul (*R.* 434-42; for the best account of this isomorphism see Keyt, 2006). But the close analogy between the tripartite city and the tripartite soul (by nature both have parts equal in number and of the same kind, 441c) that the isomorphism demands, suggests strongly, I think, that in his division of the soul Plato supposes that the human soul comes with a natural division of parts (powers or faculties, not agents, in my view) and psychic labors (functions) unique (exclusive) to each part; and further that the human soul can be educated so that its parts are matched optimally to the psychic functions of an individual to provide for, defend, and govern herself. The isomorphism itself, which is very demanding

(psychic parts must correspond to city parts and psychic functions to city functions), requires that the cryptic formula of justice in Bk. iv, that in both cases of city and soul justice obtains when “each [part] is doing its own” has to have the same interpretation; and if in the case of the city it means that each part of the city must do that city function (of the three main city functions) which it can do best (i.e. optimally), then it must mean the same thing for the soul; a soul is just when each part of it is performing that psychic function (of the corresponding three main psychic functions) which it can do best (i.e., optimally): e.g. reason must rule because it can do that better than spirit can rule, better than appetite can rule, and so on. The functional theory is in the background here, as it is also in the analogy between justice in the soul and health in the body, since bodily health was conceived in terms of bodily parts and their functions.

A third use of the functional theory, I think, can be found in Plato’s defense of justice. His explicit first defense of psychic justice by the health analogy at the end of Bk. iv, presupposes, I believe, the idea of bodily health as the well functioning of an organism. But aside from this, and also aside from his defense of justice in terms of his distinctions among pleasures in book IX, I think he has a more fundamental defense of both city and soul justice, which relies on the idea that happiness obtains when city and soul are functioning well. The functional theory itself supposes that functioning well, the fundamental normative idea in terms of which every other value in the theory is accounted by, is good for the thing that so functions: functioning well physically is good for the body; functioning well socially is good for the city; and functioning well psychologically is good for the psyche. If happiness is a good, indeed the chief and most inclusive good, then functioning well will be at least part of it. I think we can see some of this when S is made to defend the city he has been constructing against Adeimantus’ objection in the opening pages of book IV, that by depriving the guardians of property and wealth (and all the goods and enjoyments these enable us to have) S is making the guardians unhappy. S makes a reply in two parts. First, in obvious contrast to T, he says that they were trying to imagine a city so organized as to promote the happiness of the city as a whole, not just the happiness of the rulers; for, they thought, it is in such a city that they would find justice. Second, he challenges the conception of happiness Adeimantus’ objection assumes, and suggests a functional conception of happiness: happiness will be found, at least in part (the other part being psychic, not social), in citizens being “the best craftsmen in their own function” (*ergon*. 421c), and concludes that in a well ordered city “each class is to be left to the share of happiness that its nature comports.” Similarly, given the isomorphism, happiness is also to be found in each part of the soul doing well what it can do best. Each part of the city or soul doing what it can do best is simply a restatement of (the formal part) of what justice is; so justice, together with the other virtues, to be sure, since they contribute to functioning well, contributes to happiness. This is a more fundamental defense of justice, because, on the assumption that functioning well is good for a thing, the good of justice is inherent in the very nature of Platonic justice.

Finally, in book X we find some further uses of the functional theory. At 601d we

are told: “There are some three arts concerned with everything, the user’s art, the maker’s and the imitator’s. Yes. Now do not the excellence, the beauty, and the rightness of every implement, living thing, and action refer solely to the use for which each is made or by nature adapted? ... the user is the one who knows most of it by experience, and he reports to the maker the good or bad effects in use of the thing he uses.” Though *ergon* (function) is not used in this passage, notice the characterization of the use in question as “the use for which each is made or by nature adopted”, which echoes the two fold account of *ergon* in Bk. I. A bit later, at 602d-603b, we have a subdivision of psychic powers, between opining on the basis of the appearances of distance and shape and size and contrary opining on the basis of measuring, counting, and weighing; the division is made on the basis of the principle of contrariety, and it is clear in the context that measuring, counting and weighing are exclusive functions of reason. Because of these functions reason is said to be the best part of the soul; and because of these exclusive functions, ruling is said to be the function of reason: “And have not measuring and numbering and weighing proved the most gracious aids to prevent the domination of our soul of the apparently greater or less or more heavier, and to give the control to that which has reckoned and numbered or even weighed? Certainly. But this surely would be the function (*ergon*) of the part of the soul that reasons and calculates.” (602de). In this whole passage, how well a thing functions is made a basis for evaluation; exclusive functions are used to individuate parts of the soul, and the exclusive functions of reason are used as a basis for assigning to it ruling as its optimal function.

We asked earlier why Plato thought that by following the procedure his Socrates actually does follow he would discover what justice is. We answered that he thought so because he had a very general theory, taken from medicine and the productive arts, which suggested a procedure for discovering the virtue(s) of anything with function(s). In turn, we can ask why he thought that this functional theory, whose logical home is the realm of natural organs and of artifacts, would be appropriate for discovering justice, a virtue whose domain is not natural organs and artifacts but cities-states and souls. A good answer to this question is not easy to find, but its outlines may be partly in the *Phaedo*, partly in the *Republic* and partly in the *Timaeus*. In the latter we have the postulation of a cosmic teleology, according to which a non-envious and good divine craftsman created the existing physical universe to be as good as matter allows, using the Platonic forms as paradigms or patterns. The physical universe is generally the best it can be, and that is why Plato suggests in the *Phaedo* that the best explanation of why things are as they are, on a cosmic scale at any rate, is that it is best for them to be that way. When in a particular case they are not the best, as in the case of a defective eye or a rusted pruning knife, the fault is to be found in the nature of body or matter. The physicians of the day had begun to discover, Plato thought, the natural goodness of the body (i.e. health), in the natural divisions of the body, the functions the natural parts served, and the structures that enabled those parts and the whole body to function well; and this is to discover the ways in which the divine craftsman created the body so as to be as best as it can be. Similarly with the astronomers and the celestial

part of the universe; while human craftsman, such as architects and shipbuilders tried to discover and create, perhaps by imitation of the structures of the physical universe, structures that would enable their objects to be as good as possible and perform their functions as well as possible. Similarly perhaps, Plato thinks of legislators and educators as trying to discover and enhance the structures in cities and souls, which would make such cities and souls the best they can be and perform their functions as well as possible.

The Significance of Methods

Plato has T, G, and S employ different methods of reasoning about justice and he has them reach different conceptions of justice. But he does not explicitly discuss the significance of using such different methods. Would anyone of his three speakers, T, G, or S, reach the result he did reach by using the others' methods? Given what Plato has actually written, we might suppose that his answer would be no. Different methods gave him different results; so why should we suppose that the three methods might give us the same result, or that Plato would think so?

These questions are not easy to answer, partly because the distinction between methods and substantive or empirical assumptions made in applying the methods might not be all that sharp. If our question includes differences in such assumptions, then it would not be too surprising if different methods gave us different results; though logically there might still be room for the same result, since the same conclusion can be validly deduced from different premises, something Plato was probably aware of from, say, the different existing proofs of the Pythagorean theorem (For further discussion see Price, 2008).

In fact, Plato does seem to criticize the method T used, by criticizing the assumption that makes the empirical investigation of justice and its result possible: that justice is to be found only in the positive laws of each society. Thus Plato implies not only that T's super conclusion is false B a conclusion he attacks independently of method in his second argument by analogy B but also that the method is wrong headed. And we can also reasonably suppose that Plato thought one would not reach S's conception of justice by using T's method, since most societies did not practice Plato's justice, and so did not have positive laws supporting Plato's justice.

However, Plato does not explicitly criticize G's method for discovering the nature of justice by looking for a contractarian origin of justice. Nor does he consider the question whether his own conception of justice would be reached by using G's method. If we raise that question for him, very probably the answer is negative. We say very probably, at most, because what results are reached by G's method does depend on what empirical assumptions are made about men and their circumstances in the state of nature (e.g. moderate scarcity, degree of rationality, motivations and conceptions of the human good), as well as procedural assumptions, such as unanimity, and the reasoning used to make a choice.

Conceivably, Plato thought that if we supply what he thought were true or reason-

able assumptions, men in a state of nature would choose his principle of social justice over the state of nature. Certainly, his main thesis in the *Republic*, that men are better off or happier being Platonically just rather than unjust, might have led him to argue that in a state of nature individuals would rationally choose his principle of social justice over the state of nature, each on the basis of his/her happiness. He might also accept the procedural assumption of unanimity of agreement, since his social temperance implies agreement among all the citizens on who should rule and who should be ruled.

But would individuals in a state of nature rationally choose Plato's principle of social justice over the less restrictive and minimal principle G has them choose? In a state of nature, would they rationally agree to give up not only their freedom to harm others, but also the freedom to choose a career or occupation in society, the freedom to own property, or the freedom to choose a mate B all the freedoms lost in Plato's just society (at least for rulers and warriors)? Here the theory of human good embedded in G's theory of justice, in which the freedom to do as one pleases is thought to be a great good, stands in the way. If we could replace G's theory of the human good with Plato's own and attribute Plato's conception of human goods and happiness to men in a state of nature, such men might well choose Plato's principle of social justice over the state of nature (and include G's minimal agreement in their choice), each making the choice on the basis of his own good as Plato conceives it. But Plato's own theory of human good, with its radical downgrading of freedom and the goods and pleasures of ordinary men (e.g., *R.* 555-61, 582-7), seems hardly something that we can attribute to men in a state of nature; at least most men as we know them have something closer to G's conception of the human good rather than Plato's. All in all, it seems doubtful that Plato could claim that G's contractarian method would produce S's result, at least if we make reasonably realistic assumptions about the state of nature and men's capacities, conceptions of the good and desires.

So, it would seem that we could not reach S's result (Platonic justice) by using either T's empirical method or G's contractarian method.

Finally, we can raise the question whether by using G's method one would reach T' result. Suppose then that in G's state of nature the warring parties have not only the two alternatives G gives them but also T' principle: their options are to stay in a state of nature, or to agree to give up equally the freedom to harm others in exchange for equal security of not being harmed by others, or to agree to the principle that justice is the advantage of the ruling party. Now off hand it is not clear how agreeing to the third option would solve their problem in the state of nature, as it is clear that the second option would. But aside from that, there is an obvious question that each party would need to ask about the third option: in the ensuing society would I be a member of the ruling party or a subject? In a Thrasymachean society, clearly a member of the ruling party would be better off than a subject would be; justice would favor him; whereas for a subject justice would be, as T says, "the other person's good", the other person being a member of the ruling party (*R.* 343c). Now if a choosing party could not reasonably

answer this question B either because of a veil of ignorance or because of too much uncertainty even without a veil B he could well reason that he would run the risk of being on the losing side of T's justice and perhaps be no much better off than in a state of nature; whereas he would run no such risk by choosing G's compromise in which he would definitely be better off than in a state of nature. On the other hand, if a choosing party did know that he would be member of the ruling party, or that he would be a subject, he would opt for Ts' principle in the former case and for G's in the latter; thus unanimity on T's option would not be rationally possible B indeed probably only a minority would go for it. With these three options, behind a veil or under uncertainty, the rational choice would be G's justice; and with complete knowledge, there would be no unanimity on Ts' justice.

In sum, in these hypothetical cases, where we considered whether each one of the three interlocutors, T, G, and S, would have reached a different result by using the method of another, we see again the significance of methods of reasoning about what justice is. Plato contributed far more than raising this question and elaborating his own answer: he discovered, and used different methods to answer the question, and displayed how the different methods can give us different answers. A lesson still very much relevant today.

Bibliography

- Bobonich, C., *Plato's Utopia Recast*, Oxford, 1995
- Burnyeat, M., Review of *Goodness and Justice*, TLS, 2002
- Hull, D. and Ruse, M., eds., *The Philosophy of Biology*, Oxford, 1998
- Irwin, Terrence, *Plato's Ethics*, Oxford, 1995, chs. 11 & 12.
- Keyt, D., and Miller, F., "Ancient Political Thought," in *Handbook of Political Theory*, edited by G. Gaus and C. Kukathas, London, 2004
- Keyt, David, "Plato on Justice," in Benson, H. H., ed. *A Companion to Plato*, Blackwell, 2006.
- Price, Anthony, "Reasoning About Justice in Plato's *Republic*," *Philosophical Inquiry*, vol. xxx, 2008.
- Rawls, John, *A Theory of Justice*, Harvard, 1971, ch. III.
- Santas, Gerasimos, *Goodness and Justice: Plato, Aristotle, and the Moderns*, Blackwell, 2001, chs. 3 & 4.
- Santas, Gerasimos, ed., *The Blackwell Guide to Plato's Republic*, Blackwell, 2006.
- Santas, Gerasimos, *Understanding Plato's Republic*, Blackwell-Wiley, 2010.
- Adam Smith, *The Wealth of Nations*, New York, 1937
- Vlastos, Gregory, *Socrates*, Cambridge, 1991

Endnotes

- ¹ The present paper draws material from the Author's "Methods of Reasoning about Justice in Plato's *Republic*," in Santas, (2006)