Urging Righteousness and Virtue: Socrates, Gorgias and the Nature of Moral Argument

Peter G. Woolcock

In the *Gorgias* Socrates claims that it is worse to be a wrong-doer than to be the victim of wrong-doing. His adversaries, Polus and Callicles, regard this position as preposterous. In this paper, I argue that, from the viewpoint of what it is rational to urge others to do, then Polus and Callicles are the ones acting irrationally, at least when the urging takes place in a consensual, as opposed to a coercive, argument situation between autonomous and competent rational agents. They are guilty of a pragmatic practical paradox, even if what they say may be rational enough to believe when held as a theoretical view about a third party. My analysis also demonstrates that rationality favours the urging of most of the other startling judgements that Socrates endorses, with particular consideration given to the centrality of shame in showing the irrationality of urging some courses of action rather than others.

Introduction

In the *Gorgias* Socrates makes a series of assertions that strike his listeners Polus and Callicles as absurd, at least initially, and strike many of us today the same way. The man who inflicts evil, Socrates claims, is necessarily more unhappy than the man who is the innocent victim of evil (475e). The wicked man is unhappy, but the unhappiest man of all is the wicked man who does not meet with justice, rebuke and punishment (472e). Tyrants like the Macedonian dictator Archelaus who can kill whomever they please and inflict confiscation and banishment on anyone they choose are the least powerful persons in a state (466). Anyone who has done wrong will only be happy if he denounces himself and cheerfully offers himself for whatever punishment his crime deserves. If his friends or family do wrong, he will not be happy unless he is the first to accuse them (480). To cap off this list of remarkable assertions, Socrates claims that, if our enemy “has stolen a lot of money he must not pay it back, but keep it and spend it on himself and his family without regard to God or man; if he has committed crimes for which the penalty is death he must not be executed. The most desirable thing would be that he should never die, but
live forever in an immortality of crime; the next best that he should live as long as possible in that condition” (480–1).

In this paper I intend to use the debate in the Gorgias to raise, and investigate, some puzzles about the rationality of moral argument. Is it a rational enterprise? If so, what is its point? Argument about what I will call “descriptive” matters, such as scientific and historical arguments, seem to be rational in that they are an effective means for achieving the purpose we have in mind in pursuing them, namely, arriving at the truth of the matter. This is so because there actually are truths in these fields to be discovered, and arguing does seem to increase the likelihood of true answers. Does moral argument have the same rational point, or would there still be rational point to it, even if there were no descriptive moral truths? It is the principal aim of this paper to show that moral argument would have rational point regardless.

I will say that people act rationally when they are not acting in a self-defeating manner, that is, they do the thing they believe is the most effective means for achieving the purpose they have in mind, on the assumption that this is not incompatible with some other purpose that matters more to them. Were it to be so, then they would act in accord with the more important purpose. Agents will be considered irrational when they knowingly act in a self-defeating manner.

Whether or not it is rational to engage in argument about a particular practical matter depends on a number of factors. One such factor is the nature of the agents engaged in the argument. I will assume that the parties to the argument are what I will call “autonomous” and “competent” (Rawls, 2001:18–24). They are autonomous in the sense that they conceive of themselves as ends in themselves, not mere instruments for the achievement of the goals and purposes of other agents but self-authenticating initiators of goals and purposes of their own. Further, they conceive of themselves as independent of any particular goals they might come to adopt, capable of revising and changing their goals in the light of new information.

They are competent in the sense that they possess the capacities necessary to engage meaningfully in argument, both in terms of acquiring and evaluating information and in terms of acting consistently with any conclusions reached as a consequence of the argument. While the parties in the Gorgias regard each other as mistaken in certain respects none of them seems to doubt that all of them are autonomous and competent in the way just described. In fact, it is precisely because they do regard each other as autonomous and competent that they regard the enterprise of argument as worth the effort.1

A second factor to be taken into account is the kind of situation in which the parties find themselves. Two kinds of argument situations are of particular interest. A “consensual argument situation” is one where the parties lack the power to force their

1 For a useful discussion of Socrates’ dialectical method, in which an alternative justification from my “urging” account is offered for the seeming simple-mindedness of Socrates’ opponents and the seeming sophistical nature of his arguments, see Krook, 1959. Also see Dilman, 1979, where Socrates’ arguments in the Gorgias are given a Wittgensteinian interpretation.
opponents to say that they agree with them. In effect, they are in a position of more or less equal power with respect to each other. If everyone does come to an agreement then this is not because they have been intimidated or threatened into agreeing but because they have freely and willingly endorsed the agreed-upon conclusion of their own accord as a result of the argument put to them. A “coercive argument situation” is, by contrast, has, as its major premise, the conditional “If you don’t do what I tell you, then you will be punished”.

When an argument is conducted in a consensual argument situation between parties who regard each other as equally ends in themselves then, I suggest, they will largely agree on what it is rational for them to endorse in how they are to treat each other. Socrates, as he says, prefers to engage in an argument on a one-to-one basis (474) but the conditions of the consensual argument situation are such that it does not matter who the particular individuals are that participate in the argument. These conditions will push them into agreeing on a common body of judgements about character and action that it is rational to endorse publicly.

**Power and happiness**

Let us return, then, to the assertions that Polus makes about what kind of life people must lead if they are to be happy. “To listen to you, Socrates,” he says, “one might think that you wouldn’t be glad to have the opportunity of doing what you please in the state rather than not, and that you don't envy a man who can kill or confiscate or imprison at will.” Socrates asks, “Justly or unjustly, do you mean?” Polus replies, “It makes no difference; he's enviable in either case, isn't he?” (469).

There are at least two things that the parties to this argument could be using the argument to achieve. On the one hand, they could be using the argument to settle the descriptive question of whether or not a person maximises their chance of happiness by maximising their power. On the other hand, they could be using the argument to influence the other to adopt a particular attitude to the use of power in the pursuit of happiness, that is, they could be using the argument to achieve a prescriptive, rather than a descriptive purpose. All of the parties in the *Gorgias* believe that a rational agent will want to maximise his own happiness. Polus believes this of Socrates. Polus, it seems, is involved in a pragmatic practical paradox in seeking to convince Socrates of the truth of the descriptive proposition that maximising an agent’s happiness requires maximising that agent’s power. Polus wants to be free to do whatever he pleases, including killing people and confiscating their property, but he knows that Socrates, as an autonomous rational agent, will not want Polus to be free in this way, and yet he reveals to Socrates that this is what he, Polus, wants, when he must know that the inevitable effect of his letting Socrates know this is that Socrates will be better placed by this knowledge to prevent Polus doing whatever he pleases.

If Polus is not to engage in this kind of self-defeating behavior, then he cannot argue in a way that reveals that he is the kind of egoist that he actually is. In
a consensual practical argument situation, both he and Socrates will have to argue on the assumption that each of them is an autonomous rational agent of equal worth, equally deserving of being treated as ends in themselves. He will have to argue as if he accepts that Socrates’ concerns are to be given equal weight with his own, that he wants them to get equal consideration with his own. In publicly defending the view he does, then, he has acted irrationally.

However, had Polus kept the debate at the purely descriptive level rather than indicating his belief that the rational thing for him or Socrates to do was to act unjustly if they had the same power as Archelaus, if he had merely kept the discussion in the third person, using Archelaus as a theoretical case as opposed to one to be followed, then he would not have acted irrationally. For this to be clear, however, the discussion would need to have been conducted in a context where it was understood that the arguments offered were meant to explore what might be true, not to indicate where the speaker’s individual value commitments lay. In other words it needed to be the context of a purely philosophical debate in which the cases were discussed in the third person, not in the first or second person, as if they were about how third parties might act, not how Socrates and Polus themselves would act. The purely philosophical question is whether or not it is ever rational for people to use power unjustly as a means to achieving their own happiness. Socrates, however, is concerned with the practical question of whether or not one ought to be the kind of person who would endorse this as the rational thing to do, which people would only do if they had the kind of values that made it rational from their point of view. Socrates makes it very clear that this is his enterprise in his closing remarks to Callicles where he explicitly urges Callicles to follow a life of righteousness and virtue. He says, “Let us follow that way and urge others to follow it, instead of the way which you in mistaken confidence are urging upon me; it is quite worthless, Callicles” (527).

Punishment and happiness

What, then, of claims about who is the happiest — the man who commits evil and gets away with it, or the man who commits evil and is justly punished? Interestingly, Socrates has an unexpected modern ally on this topic. Osama bin Laden once said, “If a man is a real Muslim and commits a crime, he can only be happy if he is justly punished” (Fisk, 2005:24).

For our purposes, the crux of the matter is not whether the man who commits evil and gets away with it is happier than the man who commits evil and is justly punished. Rather, the issue is which of these two alternatives is it rational for a man to publicly endorse in an argument? The wrong answer will lead his listeners to conclude that argument with him is a waste of time, showing him to be the kind of person who would have no intention of abiding by any agreement they reached anyway. Suppose Polus says to Socrates, “Yes, I am the sort of person who would be happier
committing evil and getting away with it than I would be if I committed evil and were justly punished”. What kind of sense is he hoping Socrates will make of that? Assume that Polus says what he does because he thinks it will serve his own ends. What end could Polus have in mind in letting Socrates know that he would be happier committing evil and getting away with it than he would be committing evil and being justly punished for it? There might, by contrast, be some point to his telling Socrates that he sought punishment for his evil acts. In a situation where he cannot force Socrates to behave one way rather than the other, it might have the effect of getting Socrates to believe that Polus is truly repentant and, therefore, unlikely, or less likely, to perform future evil acts. Of course, he may not be telling the truth but, whether true or not, it might get Socrates to cooperate with him in ways useful for the furthering of Polus’s own ends.

However, to tell Socrates that he is the kind of person who would not be happy to be justly punished for having committed evil is to admit that the evil he commits is deliberate and intentional and likely to be repeated, and shows that he is unrepentant, whereas an active desire to be punished indicates that he did not intend the act, that it was due to weakness of will or a misunderstanding of the situation or whatever and that he wishes to offer redress for the harm he has caused. On the assumption, then, that Polus does not want to give other people reasons to limit his freedom to live his life as much as possible in accordance with his own preferences, he will not admit (in a consensual argument situation) to being the kind of person who would be happy to do evil and escape punishment.

In fact, he would want others to think that he was the kind of person who would regard a fall into temptation as a form of corruption of his soul, as a rottenness in his spirit on a par with leprosy in the body, something deserving the spiritual equivalent of being cauterised out or surgically removed, just as Socrates describes it. Again, it would be a pragmatic paradox not to do so. It is a contradiction in the sense that, with respect to his own ends, a failure to publicly declare his own evil-doing as corrupting is something he recognises would subvert his ends. In other words, such a failure to do so would be self-defeating.

We may even have an explanation here for why Socrates thinks people only commit evil acts out of ignorance. One way in which others may come to accept that you are not an evil person at heart is for them to believe that you were ignorant that what you were doing was wrong but that, now you realise it, you will ensure that you do not repeat the behaviour, that you are aghast at what you have done and ashamed not to have understood its nature.

**Inflicting evil to avoid evil**

What about the case where Polus is confronted with a choice between doing evil to others or suffering it himself? He sees them both as bad options but, if he had to make a choice, he would choose what he sees as the least bad option for himself, namely
to do evil to others. Would it be rational of him to let other people such as Socrates know this about himself?

Suppose, for example, a vicious tyrant told Polus he could choose between killing another person or having his own big toe cut off (to modify an example from Parfit, 2011:207). Unless Polus had a fanatical commitment to honesty, it would be self-defeating of Polus to say to anyone who could be that other person that, were such a situation to arise, he would choose to save his own big toe. Such a revelation would ensure that those other people never knowingly placed themselves in a situation where their welfare depended on Polus putting himself even at a minor risk. As a consequence, given that it is highly likely that some of the most important of Polus' goals will depend on other people placing themselves in such a situation, then Polus' honesty would have guaranteed that these goals would not be met. It would be a practical paradox to make public that one was the kind of person who valued himself so highly in comparison to others that he would do a major evil to others to avoid a minor evil to himself.

To avoid this paradox, Polus would need to give others the impression that he and they were of equal value. As such, he can admit that he would save his own life in preference to the life of another because his life has the same value as that of the other, but only if he could do so by fair means. He could admit, for example, that he would swim as fast as he could to be the first person to the only lifebuoy in a shipwreck but not that, were someone else to beat him to the lifebuoy, that he would pull out a knife and stab that person to death. This would be to reveal that his claims to regard other people as of equal worth with himself were entirely insincere. By contrast, were he to declare that he would intentionally let someone else (his wife, say) beat him to the lifebuoy because he valued that person more than himself, this would not generate a practical paradox.

**Corruption and happiness**

This brings us to the most outrageous of Socrates' claims, namely, his claim that, when an enemy of ours injures a third party, we should connive with that enemy to ensure that he avoids punishment, helping him to live a long life in the unhindered enjoyment of his ill-gotten gains. What Socrates is getting at here is that he, himself, is the kind of person who would find the corruption of the soul that the evil-doer undergoes as far worse than being punished and redeemed, even if the punishment was death, that he sees the punishment as the cure for a bad condition. What we have here is a particular use of the word "real" or "really". The really moral man, the one that it is in Polus's interests as a rational agent to give the impression to other people that he is, is a man who would be made totally miserable by not redeeming himself. It would be a worse punishment to him than the cleansing punishment of justice.

---

2 Thanks to an anonymous reviewer for suggesting further discussion of this issue.
Osama bin Laden uses the word “real” in his term “real Muslim” in exactly the same way. What end of Polus's would be served by his saying to Socrates that he would not be miserable in this way, that he would, in fact, revel in the opportunity to enjoy his ill-gotten gains forever? In a situation where he has more or less equal power with Socrates, where he will often need Socrates’ cooperation or, at least, not his active hostility, it would be counter-productive of him to cause Socrates to think he was an unscrupulous and ruthless exploiter of others who would not be ashamed of the harm he caused them. His interests would be best served by convincing Socrates that he was a truly moral person, and one step towards convincing Socrates of this would be to say that he regarded the corruption of his soul as a worse punishment than any he could be justly dealt, no matter how serious, even death.

It should be noted, however, that, while this is what it is rational for him to say about himself, if he is to avoid being seen as an egoist, he does not need to draw the more extreme conclusion that Socrates does, nor does rationality require him to believe what he says. Moreover, he can admit that it is a descriptive truth that there are people whose happiness is in no way diminished by the evil of their acts and the corruption of their moral nature. No obvious irrationality is involved in his admitting that not everybody is a true moral agent like Socrates and himself. Consequently, he can reject as a travesty of justice Socrates’ proposal that our enemy is most effectively punished merely by allowing the malignancy in his soul free rein.

Now, this proposal of Socrates makes sense if we accept his controversial view that people only commit evil acts out of ignorance. It certainly would be rational to say to someone who accused you of an undeniable crime that you did not realise at the time the wrongness of what you were doing. This indicates a current recognition of its wrongness and implies a resolution not to repeat it now that you know it is wrong. As such, it reassures your listeners about your status as a moral agent. However, rationality does not require you to agree that all other people who act wrongly have done so out of ignorance. Once again, there is an important difference between, on the one hand, what actions, performed either by yourself or other people, it is rational for you to prescribe publicly and, on the other hand, what it is rational for you to offer as a description of what other people endorse without thereby endorsing it yourself.

Gorgias and shame

An emotion that plays an especially important role in the Gorgias is shame (Tarnopol-sky, 2010). As it happens, much of what is explained in terms of shame in the Gorgias can be seen as an outcome of the kind of rational restrictions on what one can say that I have highlighted in this paper.

Consider the case of Gorgias himself. Socrates’ discussion with Gorgias is a preliminary to his argument with Polus about when a person can properly be said to have power. Gorgias, in effect, is boasting that oratory is a good thing because the orator is a man with a particularly useful kind of power, namely, the power to bend people
to his will with his oratory. Saying that it is a good thing amounts to saying that it is something you endorse, that you would pursue. However, it is not a rational move to announce to people on whom you might use this power that you endorse your using it in a way that is indifferent to its effects on them, that you think exploiting them in this way would be a good thing. Socrates picks up on this, getting Gorgias to admit that he would only use his power of oratory for morally acceptable purposes. Socrates plays on the fact that Gorgias would be ashamed to have people think that he would exploit them for his own selfish advantage. To avoid this shame, Gorgias finds himself, in effect, admitting that oratory gives the orator a power that, as a good man, he will rarely use. In fact, Socrates can only think of one use for it, and that is the bizarre case we came across earlier, of persuading people not to bring their enemies to justice.

By the end of the argument, his rationality requires Gorgias to say, even if he doesn't believe it, that the power he attributes to oratory is one that he will rarely use, and a power that is rarely used hardly counts as a power at all. For all practical purposes, Socrates’ argument has led Gorgias to withdraw his original assertion. However, this withdrawal should not be seen as a demonstration by Socrates that oratory does not possess the power that Gorgias first claimed it had. Clearly it does, if all that is taken into consideration is a purely descriptive account of power. Instead, what Socrates has done is to get Gorgias to withdraw his initial endorsement of this power, his contention that it was a good thing. But just because a particular power is not a good thing, it doesn't follow that it stops being a power, although Socrates often seems to think that this kind of conclusion does follow.

**Callicles and pleasure**

What, then, of Callicles? Callicles thinks that happiness is constituted by luxury and excess and licence. In order to be happy, therefore, to live as he ought, “a man should encourage his appetites to be as strong as possible instead of repressing them, and by means of his courage and intelligence to satisfy them in all their intensity by providing them with whatever they happen to desire” (492). As we noted earlier, all the parties in the *Gorgias* see his own personal happiness as the rational thing for a man to pursue. As a consequence, they regard happiness as the good, where the good is that which they would endorse as worthy of pursuit. So, if the happy life is the life of pleasure, and Callicles says that happiness is the good, then Callicles is committed to endorsing any form of life that gives the person who lives it a maximum indulgence in what they enjoy. As Socrates cunningly reveals, Callicles is not prepared to do anything of the kind. Socrates says to him, “Take the life of a catamite; is not that dreadful and shameful and wretched? Or will you dare say that such people are happy provided that they have an abundant supply of what they want?” Callicles's response clearly indicates that he regards the life of a catamite as shameful and not something he thinks is to be pursued or endorsed i.e. good. None of this shows that, at the purely descriptive level, the life of a catamite cannot be a happy one but Socrates is not operating at the
descriptive level. For him, and for Callicles, a speaker should not say that a life is a happy one unless he would regard it as a good one, and he should not regard it as a good one unless he can endorse it and commend its pursuit. Callicles cannot commend the pursuit of the way of life of a catamite because he finds it a shameful life so he is pushed into agreeing with Socrates that it is not the case that all pleasures, regardless of their nature, are good.

Even if he hadn’t regarded the life of a catamite as shameful, Callicles is in trouble trying to convince his listeners that there is no distinction between good and bad pleasures. Why would it be rational of Socrates, or anyone else, to accept that a way of life that took pleasure in their pain, or that could only prosper by the thwarting of their goals or purposes, was one that they should endorse?

Conclusion

What we have seen in our discussion of the *Gorgias* is that rational, autonomous, competent agents will find it rational to publicly endorse certain kinds of action and character when they engage in argument in a consensual argument situation. It seems that this would still be so even if there were no descriptive moral truths, for example, even if no error was involved in not believing that people were equally valuable as ends in themselves. Socrates himself seems to recognise the limits of what argument alone can achieve when he resorts to the myths of the leaky soul (493) and the naked judgement (523) to make the moral life as attractive as possible. In the situation of a tyrant like Archelaus and his subjects, what it will be rational for those with superior power to say publicly will differ from what it is rational for those with inferior power to say. Unlike scientific or historical truths, truths about what it is rational to endorse are relative to the powers of the parties engaged in the argument, as well as to the values that motivate them. Rationality in itself does not seem to require us to value other rational, autonomous, competent agents, including ourselves, as equally worthwhile ends in themselves. However, for those of us who see each other as parties to a consensual argument situation (and I assume that this includes you and me) then it would be a pragmatic practical paradox for any of us to tell the others that we do not consider them equally worthwhile ends in themselves. If this so, then, as I believe I have shown, we will also endorse almost all of the conclusions that Socrates urges upon us in the *Gorgias*. 
Bibliography

Dilman, 1979

Fisk, 2005

Krook, 1959

Parfit, 2011

Plato, 1960

Rawls, 2001

Tarnopolsky, 2010