Preferences, Prisoners and Private Information : Was Socrates Rational at his Trial?

Brishti Guha[1]

Singapore Management University

Abstract

Using concepts from game theory, political economy, law and economics and the economics of asymmetric information, we describe the economics of one of the most famous trials in history – that of the Athenian philosopher Socrates. We discuss the question of whether Socrates' actions during his trial were rational, using two different models. Our analysis sheds some light on institutional efficiency in trials that followed the classical Athenian pattern.

JEL Classification : D02, D82, K41, K42, N43

1. Introduction

In this paper we examine the economics underpinning one of the most famous trials in history – the trial of the Athenian philosopher Socrates in 399 BCE. We model the trial (which followed the pattern of other trials in classical Athens) as a sequential game and discuss issues relating to it from the viewpoint of political economy, law and economics and the economics of asymmetric information.

Our plan for the rest of this paper is as follows. In section 2 we outline the timeline of events in a typical public trial in classical Athens (of which Socrates' trial was one example) modeling the trial as a sequential game. In section 3, we provide some historical and political background on Socrates' trial as well as details on his accusers, the jury and the particulars of the trial itself. In sections 4 and 5, we consider alternative models to explain the behavior of Socrates and that of his jurors during his trial. In the process we refer to historical evidence whenever our discussion is enriched by it. We also uncover sources of inefficiency in the system and propose a more efficient mechanism. In section 6 we briefly model Socrates' behavior after his trial, in particular his refusal to escape from prison. We also discuss the relative merits of our two

models from sections 4 and 5 and debate which of them is more compatible with historical evidence and with Socrates' post-trial behavior in section 6. In section 7 we survey some related literature, in particular contrasting our approach and results with McCannon (forthcoming), and conclude.

2. Trials as Sequential Games

A significant subset of Athenian trials in the classical period were public prosecutions (graphe) which could be brought by any member of the public against a defendant[2]. Very frequently, graphe were also "timetos" which meant that there was no fixed penalty for the offense. As Socrates' trial was both a graphe and a timetos, we confine our attention to trials of this kind. Such trials followed the timeline described below.

At t = 1, accusers bring a charge against the defendant. Simultaneously, the accuser(s) also propose a punishment **p** which represents their recommended punishment in the event the defendant is found guilty. The defendant makes a speech in his own defence.[3] At the end of the period, the jury – which was drawn randomly from male adult citizens of all classes and numbered several hundreds – votes on whether the defendant was guilty or innocent.[4]

If the defendant is found innocent by a simple majority, the game stops and the defendant is acquitted.

If a simple majority finds the defendant guilty, the second round of the game commences as we move on to t = 2. The defendant then gives a second speech, proposing a counter-punishment **c**. At this point the defendant has information on the percentage of the jury who voted him guilty in the first round, and also has information on the accusers' proposed punishment **p**. At the end of the period, the jury votes between the two proposed punishments – that proposed by the accuser and that proposed in response by the defendant. Once again, simple majority voting is used to determine the outcome and this is the sentence that is implemented. Thus the outcome space **o** has 3 elements : **o = {I, p, c}** where **I** is the outcome of innocence having been established after the first round of voting.

---

[2] To discourage frivolous cases, the plaintiff had to pay the court a fine of 1,000 drachma if he failed to secure at least a fifth of jury votes in his favor (Hansen 1991, Christ 1998).
[3] Neither plaintiffs nor defendants used lawyers. Any man who would not present his own case was regarded as "having something to hide" and in fact using lawyers was illegal (Hansen 1991).
[4] We describe jury selection and composition in more detail in a later section.

3. Socrates' Trial

3.1 Historical and political background

Events in Socrates' trial followed the same timeline sketched above. His accusers – Meletus, Anytus and Lycon – brought charges against him. The charges claimed that Socrates was a man who "corrupted the young, did not believe in the gods, and created new deities". However, to understand possible factors that could have motivated Socrates' accusers, and influenced the jury, we need to look beyond these specific charges and delve into the historical and political background of Athens in Socrates' lifetime.

By the time of Socrates' birth (in 469 BC) Athens was an intensely participatory direct democracy. There was full suffrage among free adult males, the legislative assembly and the judiciary were essentially composed of very large bodies of the general public chosen annually by lot, with annual elections for a few important executive positions and full accountability to the public (Ober 2008). The decades preceding Socrates' trial, however, witnessed repeated military conflict between Athens and Sparta (the Peleponnesian Wars). There was an atmosphere of constant tension, mistrust and fear of Athenian democracy being overthrown by Spartan military rule. Socrates' student Alcibiades led a disastrous military expedition to Sicily during this period (in 415 BC) and later escaped to Sparta where he proposed to help the Spartans in defeating Athens (Linder 2002). He was also the mastermind behind the oligarchy of the Four Hundred which overthrew Athens' democratic government for a brief period in 411-410 BC. Although democracy was restored again by 410, a second student of Socrates named Critias later led a much more brutal oligarchic regime called "The Thirty Tyrants" which was imposed on Athens with Spartan help in 404-403 BC. This regime executed about 1,500 of Athens' most prominent democrats, drove 5,000 women, children and slaves into exile and confiscated the property of Athenian aristocrats. After democracy was restored in 403 following a violent resistance movement against The Thirty (this resistance was led, among others, by Anytus, who was later one of Socrates' accusers) an amnesty was declared. This meant that no one (except the tyrants themselves) could be prosecuted for perceived political crimes committed prior to 403 BC.

There is thus reason to believe that Socrates was probably viewed with suspicion by many as being the teacher of two of the most unpopular tyrants in Athenian history. The comic poet and playwright Aristophanes lampooned Socrates in his play "The Clouds" (written in 423 BC). While this play merely made fun of Socrates as an eccentric rather than depicting him as a threat to democracy, the tone changed in a later play – also by Aristophanes – called "The Birds" (417 BC). In the play, a group of pro-Spartan youth is referred to as "Socratified". As plays were widely exhibited - free of charge to the poor, who were even paid to attend them[5] (Kyriazis 2009) – depictions of Socrates as a sponsor of subversive views might well have influenced the public's perception of him.

Although the charges against Socrates had nothing to do with politics – indeed, they could not, because of the amnesty of 403 – unfavorable perceptions of his political views might have at least partly motivated his accusers and also influenced his jurors. We now turn to other possible motivations influencing his accusers, and then to an account of jury composition and selection, before describing the events of the actual trial.

3.2 Socrates' Accusers

The most powerful of Socrates' three accusers was believed to be Anytus, one of the generals who led the resistance movement against the Thirty Tyrants. Anytus' reasons for disliking Socrates were not just political, however. Socrates' habit of (usually very publicly) questioning established wisdom earned him many enemies : Plato reports an argument between Socrates and Anytus in the Meno which concludes with an enraged Anytus warning Socrates :

> "Socrates, I think you are too eager to speak evil of men: and if you will take my advice, I would recommend you to be careful".

An additional personal reason for disliking Socrates was that Socrates had a relationship with Anytus' son, and had specifically advised him against following the profession (tanning hides) that Anytus had selected for him.[6]

---

[5]*The state made it obligatory for rich citizens to sponsor theatrical performances, which it regarded as educational for the masses. It also had a fund (theorica) for compensating the poor who attended such performances, on the grounds that wages would have to be foregone in order to watch plays.*

[6] *Socrates alludes to this relationship – and the advice – in Xenophon's account of The Apology : also see Linder (2002).*

Socrates' other two accusers, Meletus, a poet, and Lycon, an orator, may have also been partly driven by personal motives, as Socrates regularly ridiculed poets and orators. Meletus was also generally regarded as Anytus' puppet.

3.3 The Jury

The composition and selection of jurors are of interest for a few different reasons. First of all, we will deal with the jury's preconceived notions about Socrates, and their preferences for punishment, in the models that we present in later sections. Therefore, it is pertinent to enquire whether the jury's preferences corresponded to the average citizen's. Secondly, we would want to assess the likelihood of a juror's being bribed. Thirdly, we may wonder whether the jurors indeed were in a position to express their true preferences or judgments while voting, or whether peer pressure (ie a fear of being blamed by other jurors with a different opinion, or by the public) colored their decisions. Finally, we would also like to assess whether jurors were prone to moral hazard – for example, trying to drag on a case, creating a "litigation explosion", or preferring sentences of fines for monetary gains. Luckily, we have enough information to address all of these questions.

All Athenian male citizens above 30 who were not state debtors were eligible for jury service. From all the applicants, 6000 were chosen by lot each year to serve as jurors. Pay for jury duty had been introduced by Pericles, and by Socrates' time jurors were paid 3 obols for a typical day at court. 3 obols, or half a drachma, was however lower than a skilled worker's daily wage of 1-1.5 drachma, though according to Hansen (1991) it was still sufficient to feed a small family. Therefore, those who volunteered for jury duty were very often the poor, or the relatively infirm or aged (Hansen 1991)[7]. In general, therefore the jury very much represented the masses, and factors which impacted the notions of the general public or the average citizen would have influenced them, as well.

We have seen that the selection of jurors was random and jurors changed frequently. Moreover, by 403 BC, an elaborate daily procedure of allotting the day's jurors to different courts was in place (Hansen 1991). Therefore, before the day of the trial no one could predict which juror would be judging a particular case. This, combined with the very large number of

---

[7] *Hansen quotes Isocrates as saying "It must pierce every right-thinking person to the heart to see citizens undergoing a selection by lot that determines whether they get their daily bread or not".*

jurors for each case (there were 501 in Socrates' case, and usually never less than 200 for any particular case) made bribery very difficult.[8]

Voting by the jury was always by secret ballot (through a fascinating procedure described in detail in Hansen (1991)). At the end of the procedure, votes were visible and could be counted, but remained anonymous. Therefore, jurors were not influenced by fears of reprisal or unpopularity while casting their votes : they could not be held personally accountable for a particular outcome.

Finally, jurors would not have personally benefited from a sentence imposing a fine on a defendant. This was because all such fines went to the state as revenue (Kyriazis 2009). Jurors also did not have the opportunity to drag on a particular case, as Athenian law required that each case be settled over one working day (nine and a half hours) and was very strict on this point (a juror was always appointed to keep a watch on the clock and time the accuser's and the defendant's speeches). While Athenians did seem to be an unusually litigious people (Christ, 1998) the jury did not seem to have played any particular role in bringing this about, though they undoubtedly benefited from the consequent high demand for jury services. However, frivolous lawsuits were usually deterred because of the practice of fining a plaintiff who failed to secure at least a fifth of jury votes (such a fine would also go to the state).

We now turn to the events during Socrates' trial.

3.4 The Events during the Trial

Socrates' accusers suggested that he deserved the death penalty in the event of conviction. We know this because Socrates refers to it during his (first round) speech in his defence. Socrates' *speech* was recorded by his contemporaries and disciples, Plato and Xenophon (it, along with two other speeches Socrates made during his trial, form "The Apology", included in Plato's *Dialogues*).   After this speech, the jury, which numbered 501, voted Socrates guilty by a relatively narrow margin of 281 to 220. In the "second round", Socrates responded to this by giving a second speech with his proposed counter-punishment. We know that he was aware of the breakdown of votes during the first round, for he says in his second speech

"There are many reasons why I am not grieved, O men of Athens, at the vote of condemnation. I expected it, and am only surprised that the votes are so nearly equal; for I had thought that the

---

[8] *According to Hansen these elaborate precautions were partly a reaction to occasional bribery scandals in the past (including one involving Anytus, Socrates' accuser, in about 411).*

majority against me would have been far larger; but now, had thirty votes gone over to the other side, I should have been acquitted."

Socrates' behavior during his second speech is of interest. He first states that, rather than being punished, he should be awarded a lifetime of free meals at the Prytaneum, a building which housed members of the Council – an honor reserved for athletes and others of great prominence. It shortly becomes clear, though, that this is a rhetorical statement, for Socrates later proceeds to propose a fine – initially a very low one of 100 drachma, about one-fifth of his own (modest) net worth – which was increased immediately to 3,000 drachma (with his supporters' backing) as his counter-punishment **c.**

Following this, there is a second round of voting, and the death penalty **p** is chosen over Socrates' proposed fine **c** by a large margin of 361:140.

4. Issues in Socrates' Trial : Jury Preferences

From the account of Socrates' trial, we may be tempted to speculate about his rationality. In other words, was Socrates' proposed counter-punishment **c** a rational response designed to avoid the death penalty (an objective in which it failed) or was it necessarily a violation of rationality? We examine this issue by building two different models in this section and in section 5.

4.1 Preferences, Lack of Common Knowledge, and Rationality

We must first model the jury's preferences. As we have seen, jury size was large, and it is reasonable to assume heterogeneity in jury preferences. After the first speech by the defendant, we assume that each member of the jury has a subjective opinion **s** regarding the extent of the defendant's guilt or innocence. The opinion is not a binary variable – it can be scaled (between 0 and 1 for convenience, where 0 represents the opinion of the jury member who believes most strongly in the defendant's innocence, and 1 being the opinion of the jury member most strongly convinced of guilt). The density of **s** is given by **f(s),** with cdf **F(s).** The opinions are a product both of the jury members' pre-conceived notions and of any additional information received during the process of the first round of the trial (from the two parties' speeches).

We may, alternatively, think of each juror drawing a realization s from a unit interval with density f(s) at the end of the first round with "nature" determining the realization, or the

juror's "type". The previous section sheds some light on why heterogeneity in s is plausible. While many people must have had negative associations about Socrates due to political reasons or due to the way he was depicted in plays, it is also likely that many knew about Anytus' personal grudges against him (Anytus and Socrates were both well-known public figures). A number of people probably also respected Socrates for his intellectual prowess and skill in debate. Thus, even prior to the trial, people might have differed in their preconceptions depending on how much weight they assigned to various bits of prior knowledge about Socrates and his accusers. In addition, s would also be influenced by Socrates' defence speech in which he not only attempted to address the actual charges but also spoke about his refusal to obey unjust commands issued by the Thirty Tyrants, and this speech would have affected different jurors differently.

The defendant (in this case Socrates) however has no way of knowing the exact form of **f,** even though he can observe the breakdown of first-round votes between his guilt and innocence. This breakdown of votes, meanwhile, would be influenced not only by **f(s)** but also by **p**. The jury was aware of the accuser's proposed punishment in the event of a first-round verdict of guilt. We will later return to the probable impact of **p** on the verdict.

After the second round of the trial, each jury member has a preference **P** regarding the scale of punishment that the defendant deserves. These punishments are rankable on a continuous scale that is common for all jurors: different jurors have different preferences over these punishments. These preferences regarding punishments have a density function **g(P(s, p, c, e))** and cdf **G(P(s, p, c, e))** where **e** is extra information that the jury may pick up during the second speech by the defendant. The accusers' proposed punishment and the counter-punishment proposed by the defendant also impact the jury's thought processes. Moreover, for any two individuals i and j from the jury,

$$s_i > s_j \rightarrow P_i(s_i, p, c, e) > P_j(s_j, p, c, e) \qquad \textbf{(1)}$$

An individual who has a stronger conviction about the defendant's guilt after the first period will also prefer a stronger punishment in the second period.

Again, the defendant cannot know the exact form of **g**. He is aware of **p** and of the percentage of the jury that voted him guilty in the first round. He has to decide on **c**. In this model, we also assume that he cannot make a rational decision on **e**, for while the contents of his speech are important elements of "extra information" for the second round votes, he cannot

predict whether a certain remark will generate a positive or negative reaction. (In our second model we drop this assumption and allow the defendant to deliberately choose **e** using it to convey a "signal" to the jury).

If the defendant were rational in the sense of wishing to avoid the accusers' proposed penalty **p,** a *sufficient* condition for him to be able to do so would be if he could propose **c < p** such that

$$G(c) > 1/2 \tag{2}$$

If the defendant knew **G**, *and* picked the lowest value of **c** that satisfies (2), he would be assured of having his proposal accepted. However, he does not know the exact form of **G**, so it might be difficult to guess about what values of **c** would satisfy (2). We prove below that if the defendant has certain beliefs about the distributions of **F** and **G**, he will always propose the lightest possible punishment for himself.

Proposition 1 : If the defendant believes that **s** and **P(s, p, c, e)** are both symmetrically distributed, then subject to the accuser's demanding the maximum punishment, he proposes the lightest possible counter-punishment. This is not dependant on the fraction of votes in favor of his innocence in the first round.

Proof : Suppose a defendant is informed of the outcome of first-round voting and learns that x percent of the jury has voted in favor of his innocence, where x < 1/2 so that a second round is guaranteed. Let p̲ be the lightest possible punishment acceptable to any member of the jury, and assume that **p**, the accuser's proposed punishment, is the heaviest possible punishment. As defendants do not make decisions on **e**, they only have to choose **c.** A defendant who picks **c** = p̲ expects that he will get the support of half the jury as he expects their preferences to be symmetrically distributed. Therefore he chooses the lightest possible punishment. Moreover, this strategy is independent of the precise value of x. **QED**

Corollary 1 : If G is symmetric and if the defendant also knows its form, he would choose the lightest **c** such that G(**p**) - G(**c**) < 1/2 or G(**c**) > 1/2 - G(**p**).

From Proposition 1, one explanation for Socrates' choosing a light counter-punishment may be that he had mistaken beliefs about the distribution of preferences (both initial opinions regarding the extent of his guilt and second-round preferences regarding punishment). If he had,

for example, believed the jurors to have more or less symmetrically distributed preferences[9], he could quite rationally have expected to have his counter-punishment accepted. That it was not indicates that a large proportion of those who believed in his guilt strongly favored severe penalties. While Corollary 1 shows how a counter-punishment could have been chosen with perfect information, this was actually never feasible (for Socrates or for any other defendants) due to lack of common knowledge of jury preferences.

The assumption that defendants lacked knowledge of the distribution of jurors' second-round preferences among punishments is easy to defend. The trial took place in one working day, and no consultation among the jurors was permitted. Moreover, the defendant would have no way of predicting which jurors would be chosen to try his case. Therefore, he would not even have an opportunity of securing information on them prior to his trial, much less would he be able to guess their opinion on punishments during the second round of the trial.

Having discussed one possible explanation for Socrates' proposal of a light counter-punishment, we now ask if this is compatible with the punishment **p** demanded by the accusers. The accusers demanded the death penalty subject to conviction, in the first round of the trial. By proposing such a heavy penalty they ran the risk of their penalty being rejected by the jury subject to the trial reaching the second round. What, then, explains their behavior? There are two possible reasons for this. First, the accusers knew that Socrates would not know the exact distribution of jury preferences. Therefore, there was a chance that the counter punishment he would propose would be too light to be preferred to the maximal penalty by a majority of the jurors. Secondly, and possibly even more importantly, the accusers would know that their choice of **p** would send the jury a signal about the strength of their conviction of Socrates' guilt. This could then affect the *first-round* outcome of the trial – whether Socrates was convicted at all, or acquitted. Indeed, if the accusers had suggested a lighter punishment like exile, disenfranchisement or imprisonment (which were all common punishments in Athens, see Allen (2000)), many of the jury might have inferred that the accusers were not convinced of the extent of Socrates' guilt. This might easily have led to acquittal. Even in the best case scenario (from the accusers' point of view) the punishment imposed would not be harsh enough to satisfy them (the accusers).

---

[9] For instance, a uniform distribution, or a normal distribution.

There is one more striking factor in the results of the second round's voting in Socrates' trial. The large majority which voted for the death penalty far exceeds the majority by which Socrates was proclaimed guilty in the first round. From this we may infer that a large number of those who had voted for his innocence initially ended up voting for the death penalty subsequently. We conjecture that the reason for this is at least partly related to **e** – extra information conveyed to them by the second round speech by the defendant. As hypothesized above, defendants may not regard **e** as a choice variable, because they have no idea of the direction or magnitude of influence of a particular remark made in the course of their speech. However, in Socrates' case it seems likely that his rhetorical remarks about deserving to be rewarded by free meals conveyed a strong negative signal, "converting" those of the jury who had only been mildly favorable earlier. In the model in section 5 we allow Socrates to choose **e** deliberately.

4. 2 An Inefficiency

In the model above, the defendant's ignorance about the distribution of jury preferences may lead him to propose a counter-punishment that is rejected, even if he is fully rational in the sense of wanting to avoid the accuser's proposed punishment. Two sources of inefficiency combine : first, the punishment space is restricted to {**p, c**}, and secondly, the jury's preferred punishments do not become public.

An alternative mechanism that would have removed this inefficiency would involve *each* jury member revealing his preferred punishment (instead of only the accuser proposing one). An arbitrator could then find the median punishment and use it as the defendant's sentence. The defendant would weakly prefer this to the accuser's proposed punishment, and the majority of the jury would prefer the median punishment to an extreme punishment. Such a mechanism could have saved Socrates. However, his jury were constrained to vote between only two punishments, rather than state their personal preferences with regard to punishments; and he himself was also ignorant about the jury's preference distribution while proposing **c.**

We note in passing that a mechanism such as that suggested in the preceding paragraph, with each jury member revealing his preferred punishment, would not imply a violation of anonymity. No verbal consultation would be needed: the preferred punishment could simply be written down anonymously using a system of secret ballot similar to that which the jury used to

vote between **p** and **c.** While this would require them to write down their preferred punishment, this would have been feasible as according to Ober (1989), literacy was universal among the Athenians of the age.

5. An alternative model : Private Information, Signaling and Nature

Our model of jury preferences in section 4 depicts Socrates as a rational individual who wished to save his life but was unable to do so. We now describe an alternative model. Socrates is still rational, but the difference is that he does not wish to avoid the accuser's proposal **p.**

Why might this happen? We argue that this could be likely if Socrates had certain private information about himself. For example, he could have known something about the state of his health which would have led him to expect to live for a very short time (later in this section we discuss some evidence that Socrates rationally preferred to receive the death penalty, without appearing to wish to do so, and suggestive evidence that he may also have received some private information about himself at a critical juncture). For the time being, we lay out the sequence of moves in such a game.

1. Prior to the trial, Nature chooses the jurors' "types" (their realizations of **s**) as in the previous model. However now Nature also determines Socrates' type (whether or not he has a terminal illness). Each juror knows only his own type, and Socrates knows whether he is terminally ill, but does not know the jury's types.

2. The trial then begins, with the accuser proposing **p**, the jury hearing evidence on both sides and making their first-round decision. Socrates observes x, the proportion of votes in his favor.

3. If the trial moves on to the second round, Socrates, proposes **c** in a speech which also conveys extra information **e.** Socrates chooses both **e** and **c** strategically. At the end of the round the jury chooses between **p** and **c.**

What would happen if Socrates received private information in step 1 that he was terminally ill? In this case, Socrates' utility from being acquitted would incorporate some utility (say B) from a very brief addition to his life span. However, it would also incorporate a disutility S from suffering. Therefore, we have $U_S (I) = B - S$, where the subscript S refers to Socrates, and I is

the outcome of being judged innocent. How would this compare with Socrates' utility from being judged guilty and sentenced either to **p** or to **c**? For any **c** which did not involve death, Socrates would get a utility of U_S (**c**) = B – S – **c**. Thus, he would have to suffer the ravages of disease while gaining a short addition to his lifespan, but would also suffer additional disutility – relative to acquittal – in proportion to the severity of the counter-punishment (for example, he might suffer disutility from having to pay a fine, or from being imprisoned). However, given a death penalty **p**, Socrates would have a slightly shorter life but escape the ravages of disease. In addition, he would gain a perhaps considerable utility M from dying a martyr's death. As we will see later from Socrates' conversations with his disciples, Socrates believed that an unjust death penalty would greatly enhance his fame and that it would guarantee his reputation among future generations. Thus we have U_S (**p**) = M.

We can now see that if M > B - S, Socrates' most preferred outcome would be to get the death penalty: however, he must be viewed as putting up a fight. This was essential for him to be regarded as a martyr by contemporaries and future generations. This explains why Socrates felt the need to suggest a counter-penalty at all, instead of simply agreeing to the death penalty proposed by his accusers. It also explains his choice of **e** and of a light **c.** Socrates, after observing x, realized that many jurors had voted in favor of his innocence. There was a very real danger that if Socrates proposed a moderate to heavy counter-penalty, such as penal imprisonment, or exile, the majority would have voted for that rather than for the death penalty. Socrates, therefore, took two precautions. First, he deliberately chose **e** to convey a very negative impression of himself to the jury. While the jury were used to defendants pleading for mercy, Socrates behaved arrogantly and suggested that he deserved to be rewarded and honored for his services to the state. This would have adversely affected the jury's second round preference distribution among punishments, skewing them in favor of a heavy penalty. Secondly, Socrates proposed a very light **c** (recall that initially he had only proposed a fine of 100 drachma, although later his supporters, who did not necessarily share either his preferences or his private information, increased the amount) to further reduce the danger that this counter-punishment would be acceptable to a majority of jurors.

Thus, this model would explain Socrates' inflammatory remarks to the jury as well as the lightness of his proposed counter-punishment. It also explains why he bothered to put up a

defence for himself, and why he did not simply accept the proposed death penalty, even if a martyr's death was his preferred outcome.

However, we have not yet considered whether the jury would be swayed by their expectations regarding Socrates' possible irrationality, or his state of health. In particular, would Socrates benefit by deliberately signaling to the jury that he was irrational (for instance through his choice of inflammatory remarks **e**), even though he was not?

We argue that although Socrates used **e** to convey a negative signal, this was meant to create an unfavorable impression of him and bias the jury in favor of the death penalty, rather than to create the impression that Socrates himself was irrational. This is because the insanity defence is a concept in modern but not in ancient law. The jurors in Socrates' era were merely given instructions, in their jurors' oath, to uphold the law when it was clear and to use their subjective judgment where it was not. Wood (2009) shows that there is no evidence whatsoever that the insanity defence was ever used in classical Athenian trials: therefore, even if the jury suspected Socrates of being irrational, this would not move them to clemency (it might even have made them fear his influence the more). In addition, we feel that given Socrates' reputation as a scholar and teacher it is very unlikely that his remarks would convey a signal of insanity, although they might well signal arrogance and lack of repentance.

We now review some evidence to suggest that Socrates did indeed feel that he lost little, and gained something, by being sentenced to death. Immediately after hearing his sentence, he says

> "Not much time will be gained, O Athenians, in return for the evil name which you will get from the detractors of the city, who will say that you killed Socrates, a wise man; for they will call me wise even although I am not wise when they want to reproach you. If you had waited a little while, your desire would have been fulfilled in the course of nature."

In Socrates' opinion, the death sentence would ensure that posterity denounced the Athenian government's actions as evil. He also believed that a martyr's death would enhance his own reputation among future generations. For these reasons, he might have rationally preferred dying from a death sentence to living for a very short time.

Xenophon, another contemporary and disciple of Socrates, also claimed that Socrates found the death sentence preferable to suffering the ravages of old age and disease. He records a conversation between Socrates and his friend Hermogenes prior to Socrates' trial. Hermogenes

urges Socrates to plan a good defence. Socrates acknowledges that until recently both he and his supporters had been trying to think of ways by which Socrates might be acquitted, but indicates that his views had recently changed :

"It was with good reason that the gods opposed my studying up my speech at the time when we held that by fair means or foul we must find some plea that would affect my acquittal. For if I had achieved this end, it is clear that now instead of passing out of life, I should merely have provided for dying in the throes of illness or vexed by old age, the sink into which all distresses flow, unrelieved by any joy".

This suggests that while Socrates might initially have preferred to be acquitted, or to receive as light a punishment as possible (as in model 1), his knowledge of his state of health at a juncture prior to his trial led him to revise his preferences and act according to model 2. Therefore, in the presence of private information about his own condition, it becomes plausible that Socrates actually did not wish to avoid the death sentence. In this event, his remarks made during the beginning of his second round speech about deserving a reward rather than a punishment may not reflect an ignorance of the effect of this remark on his listeners (as in our earlier model where we had assumed that defendants cannot predict the effects of their remarks). On the contrary, he may well have anticipated a negative reaction and chosen to influence **e** in such a way as to antagonize jury members and induce a vote favoring the death penalty.

6. Socrates' Behavior After His Trial: our two models evaluated

So far we have focused on Socrates' actions during his trial. However, his post-trial behavior is also of interest. In this section we discuss this and evaluate its compatibility with the two models developed above.

6.1 After the Trial

After his trial, Socrates was put in prison awaiting his death sentence. Plato's "Crito" records a conversation between Socrates and his friend and disciple Crito, who came to see him in prison. Warning Socrates that his sentence might be carried out the next day, Crito urges Socrates to escape from prison, explaining that he and other well-wishers had arranged to bribe the prison guards and would help Socrates escape at night. Crito suggests that Socrates flee to Crito's

friends in Thessaly. However, Socrates refuses this option and gives a number of reasons for his choice.

First, Socrates feels that escaping would endanger his friends and supporters, who would almost certainly be caught and punished with either disenfranchisement, exile or at the very least substantial loss of property. Secondly, he is concerned that escape would signal to present and future generations that he had no respect for the law and must, by extension, have been truly guilty of the crimes of corrupting the young, and of impiety, with which he was charged. Far from achieving a martyr's death, he would gain infamy. Thirdly, he appears to consider Thessaly exceptionally lawless and says that the pain of a life among such men would not be worth the very brief extension of life that escape would bring him. Lastly, in response to Crito's argument that he should try to escape death in order to be able to take care of his minor children, he points out that he would not be able to help them by escaping. He could not take them to Thessaly with him – thereby depriving them of Athenian citizenship – and therefore would have to depend on well-wishers to help bring them up, which he would also have to do in case he stayed and accepted his death sentence.

6.2 Compatibility with Models 1 and 2

Does the behavior described above favor either one of the two models we have developed earlier? Socrates' utility from escaping can be expressed as $U_S(E) = B - \alpha S - M$ where $\alpha > 1$. Socrates would gain a brief period of life but his sufferings would be magnified both on account of having endangered his friends and on account of having to live among lawless men. Instead of gaining a martyr's utility he now receives a corresponding disutility of $-M$: the wages of breaking the law and escaping a death sentence are a damaged reputation.

If we consider model 2, we see that Socrates' utility from staying put and accepting his death sentence is M. Now recall that $M > B - S > B - \alpha S$ (as $\alpha > 1$). Therefore we also have $2M > B - \alpha S$, or $M > B - \alpha S - M$: Socrates would prefer the death sentence to escape. Therefore, his pattern of behavior is consistent with model 2.

What if model 1 held, that is Socrates had no private information about being ill, and had in fact wanted to save his life but failed? We argue here that Socrates' most preferred option would be to have been acquitted: next, he would prefer being let off with a light counter-punishment while he would get least utility from death. However, even in this case, death from an unjust punishment would give him the utility of a martyr, and all available evidence indicates

that this was of importance to him. Thus, we might have $U_S (\mathbf{p}) = M$, $U_S (\mathbf{I}) = L$, $U_S (\mathbf{c}) = L – \mathbf{c}$, where L indicates the value of life to Socrates, and $L > L – \mathbf{c} > M$ (for the value of $\mathbf{c}$ chosen by Socrates). However, his ignorance of jury preferences led him to be sentenced to a death penalty. Even in this case, he might rationally prefer not to escape. Escape would give him a negative reputation, as before, resulting in a disutility of $–M$: in addition, although he would get a utility L from life, this would be reduced by a disutility D from endangering his friends and living among lawless men. Thus we would have $U_S (E) = L – M – D$. It is easy to verify that for M in the range $\{(L – D)/2, L – \mathbf{c}\}$, Socrates would not escape from prison, even when model 1 holds.

Thus while Socrates' behavior after his trial is of independent interest, it does not provide conclusive proof on which of the two models is a better explanation of Socrates' behavior. We have sketched the economics underlying Socrates' trial using two different models – one based on the defendant's ignorance about the distribution of jury preferences, and the other based on the defendant's possessing private information about himself - mainly because neither alternative can be accepted as certain, while exploring both gives us a rich picture of different economic concepts at work. However, in view of some evidence mentioned at the end of section 5 - given Xenophon's accounts and Socrates' interest in exposing the evils of the Athenian government – as well as his repeated references to fearing a life of disease, we feel that the second model is somewhat more likely than the first.

Interestingly, if Socrates' preferred outcome had been to die a martyr's death, his death achieved his objectives. According to the third century historian Diogenes Laertius, Athenians felt such remorse after Socrates' death that they shut up the training grounds and gymnasia and later honored Socrates with a bronze statue. In a sense, his death served to immortalize him.

7. Comparison with related literature and Conclusion

While we have used, and cited, a plethora of sources by historians and legal scholars regarding ancient Athens, her legal system and Socrates' trial, there are very few papers in economics that mention Socrates' trial. One such paper is by Flanagan (1991). In that paper, the author mentions the trial as an example of final-offer selection (a concept used in dispute resolution). However, unlike in our paper there is no economic analysis of Socrates' trial.

Another paper which contains an interesting analysis of Socrates' trial is McCannon (forthcoming). Below, we contrast our approach and results with his.

McCannon's focus is on showing that the classical Athenian pattern of public trials had several desirable properties, and in analyzing the impact on deterrence relative to the modern system where unanimity is often required in criminal charges. In contrast to our approach, he assumes that the punishments preferred by all the jurors are common knowledge and that these preferences remain unchanged between the two rounds of the trial. As a consequence, he finds that convictions always occur, with the accusers suggesting a penalty corresponding to the median juror's preferred punishment, which is accepted. He also assumes that jurors are motivated to find a punishment to "fit the crime".[10]

In contrast to McCannon, our focus is on exploring the rationality of Socrates and of discussing the institutional inefficiencies in the system. An important reason for the difference in results is that we drop the common knowledge assumption, and also allow jury preferences to change between the two rounds – for instance, in response to **e**. Because of this, our framework is compatible with accusers suggesting even the maximal, instead of the median, punishment. Also, unlike McCannon's model, ours does not imply that convictions always took place in equilibrium. In fact, we are aware that acquittals occurred in several trials of the period. For instance, Ober (2008) mentions a trial in which a prominent citizen, Lycurgus, accuses an ordinary man, Leocrates, of treason in 330 BC. This resulted in a public trial: however in spite of a persuasive speech by Lycurgus, the jury acquitted Leocrates for lack of proof that he had broken the law. In another account, Hansen (1991) mentions a public trial against three generals, Timotheos, Iphikrates and Menestheus, in 356-5 after the defeat of Athens at Embata during the Social War. The jury acquitted Menestheus, dismissed Iphikrates and fined Timotheos. Thus, our framework has the advantage of being able to accommodate such cases. The difference in our assumptions regarding common knowledge also allows us to elaborate on inefficiencies in the system and to suggest an alternative.

---

[10] While this is regarded as a desirable feature in modern jurisprudence, there is no evidence that it was an objective in ancient Athens. As we have mentioned, the juror's oath simply exhorted jurors to uphold the law and did not give specific guidelines with regard to appropriate punishments.

References

Allen, Danielle S (2000) : *The world of Prometheus : the politics of punishing in democratic Athens.* Princeton University Press. Princeton, NJ, USA and Woodstock, Oxfordshire, UK.

Christ, M.R (1998) : *The Litigious Athenian*. Johns Hopkins University Press, Baltimore.

Diogenes Laertius "Life of Socrates", in "The Lives and Opinions of Eminent Philosophers", translated by C.D Yonge, http://classicpersuasion.org/pw/diogenes/dlsocrates.htm.

Flanagan, Robert J. (1991) "Socrates Confronts Final-Offer Selection", *Industrial Relations* 30(1) : 163-164.

Hansen, M.H (1991) : *The Athenian Democracy in the Age of Demosthenes : Structure, Principles and Ideology.*. Blackwell, Oxford, UK and Cambridge, USA.

Hickie, W.J (1853) : *The Comedies of Aristophanes : A New and Literal Translation.* Volume 1. Henry G Bohn, London.

Kyriazis, N (2009) "Financing the Athenian state : public choice in the age of Demosthenes", *European Journal of Law and Economics*, 27:109-127.

Linder, D (2002) "The Trial of Socrates", SSRN Working Paper.

McCannon, Bryan C "The Median Juror and the Trial of Socrates", *European Journal of Political Economy*, forthcoming.

Ober, J (1989) : *Mass and elite in democratic Athens : rhetoric, ideology and the power of the people*. Princeton University Press, Princeton.

------, (2008) : *Democracy and Knowledge : Innovation and Learning in Classical Athens*. Princeton University Press, Princeton and Oxford.

Plato "Apology : an account of Socrates' Defense against the Charge of Impiety and Corrupting the Young", in Dialogues of Plato. New York: Pocket Books, 1951, pp 1-27.

-----, "The Dialogues of Plato" translated by Benjamin Jowett. Harper & Bros, 1914.

----, "Apology (the Defence of Socrates)", in Great Dialogues of Plato (translated by Eric Herbert Warmington, Philip G Rouse). Baker and Taylor Books, 1999, pp 423-460.

----, "Crito" translated by Benjamin Jowett, available at http://classics.mit.edu/Plato/crito.html.

----, "Meno" translated by Benjamin Jowett, available at http://classics.mit.edu/Plato/meno.html

Todd, S.C (1993) : *The Shape of Athenian Law.* Clarendon Press, Oxford.

Xenophon "The Apology", English translation by Henry Graham Dakyns, available at http://www.gutenberg.org/etext/1171.

Wood, Debra (2009) "Ancient Origins – or otherwise – of the Insanity Defence", *Psychiatry, Psychology and Law*, 16, S145-S151.