

Abstract

A model of the famous Trial of Socrates is presented. Two important features of the trial are, first, both the prosecutor and the defendant propose sanctions and, second, a jury conducts a simple majority vote over the two alternatives. This is shown to have good normative properties. First, a median-juror result is derived. This sanction is a Condorcet Winner among all possible sanctions. Second, the imposed sanction responds to both the harm of the act and the uncertainty of guilt. Third, under mild conditions, the expected sanction that arises should deter resulting in fewer crimes being committed.

Keywords: crime, deterrence, sanction, Socrates, trials

The Trial of Socrates¹

Bryan C. McCannon
Wake Forest University
mccannbc@wfu.edu
(336) 758-5130

April 2008

¹I would like to thank Robert Tollison for a helpful discussion and Nicholas Smith for his many valuable suggestions and comments on Classical Athenian justice.

1 Introduction

In 399 B.C. one of the most famous trials took place. The philosopher Socrates was accused of “corrupting the youth” and “not believing in the gods”. For these two offenses Socrates was sentenced to die by drinking a poison, hemlock. While the writings of Plato and Xenophon on this trial serve as a foundation of the study of Philosophy, the trial used to prosecute and convict Socrates has received no attention in Law and Economics. The goal here is to formally model the Trial of Socrates and argue that the system used by the Athenians at this time has some good normative properties that, if implemented, would result in an improvement in the way the accused are punished.¹

In the Trial of Socrates a prosecutor brings charges against a defendant and proposes a sanction. The jury observes evidence and listens to arguments presented by both the prosecutor and defendant and then votes on whether to convict or acquit. If a simple majority of the jury votes to convict, then the defendant offers a sanction as a counterproposal. The jury then casts a second vote for either the prosecutor’s sanction or the defendant’s sanction. The proposal that receives a majority of the votes is the implemented punishment. It is shown that the Trial of Socrates results in a median-juror result. The members of the jury must assess both how much harm was caused by an act and whether or not the individual on trial actually committed the offense. Consequently, each juror calculates an expected harm and prefers the outcome to be as close to its expected harm as possible. The competition between the prosecutor and the defendant results in the median-juror’s assessed harm being the sanction imposed.

The median juror’s ideal sanction has appealing properties. First, it is the Condorcet Winner. If the members of the jury held a pairwise vote against any other sanction, the median-juror’s sanction would win. The members of the jury have been presented the evidence on both the harm caused and guilt of the defendant and are, thus, in a good position to impose the sanction that “fits the crime”. Second, a criticism of current jury trials is that the sanction imposed on the convicted does not take into consideration the quality of the evidence against the defendant (Andreoni, 1991; Lando, 2005; Miceli,

¹This is not the first time the classical Greeks are used to analyze economic institutions. Farquharson (1969) and Riker (1986) study Pliny the Younger’s proposed voting procedures to illustrate the possibility of non-implementation of the Condorcet winner when there are three alternatives.

1990; Rubinfeld and Sappington, 1987). Jurors are asked to vote only on whether to convict or acquit. This results in a threshold level of certainty where a juror who is less certain about the guilt votes to acquit, even if the probability of guilt is rather high. Conversely, a juror may vote to convict even if she is rather uncertain as to whether the defendant is actually guilty (Andreoni, 1991; Connolly, 1987; Miceli, 1990). The sanction imposed is determined by the median-juror's assessment of the expected harm caused by the defendant and, thus, the more certain she is of the defendant's guilt the more severe the sanction. Similarly, the greater the harm that is caused by the offense the greater the sanction. Thus, in the Trial of Socrates the imposed punishment adjusts to both the harm of the act and the quality of evidence and certainty of guilt of the defendant. Finally, an individual who is deciding whether to break a law must now be concerned with the preferred sanction of the median juror, rather than the minimum juror - the one least willing to vote to convict. Under mild conditions the expected sanction is higher in such a trial than one that requires unanimous consent and imposes a fixed sanction.

A brief overview of the Trial of Socrates and Athenian justice is provided in Section 2. Section 3 presents a simple model of the trial and Section 4 solves for the equilibrium and considers extensions to the environment. Section 5 concludes.

2 A Brief Description of Athenian Justice

I now turn to a brief description of Athenian justice in 399 B.C. and a depiction of what is known of Socrates' trial. Interested readers are encouraged to see Brickhouse and Smith (1989) and Stone (1989) for a discussion of the specifics of Socrates' trial and Glotz (1929), MacDowell (1978), and Todd (1993) for details of Athenian political and judicial institutions. Millett (2005) discusses the accuracy of the accounts of the trial arguing that since there were many jurors and witnesses the accounts must be accurate, given the Athenian standard of truth. Therefore, I treat the accounts of the trial presented by Plato and Xenophon as factually accurate descriptions of the Trial of Socrates.

Athenian justice is characterized by private prosecution of crimes. Private citizens brought charges against those they believed broke the law. For example, in courts of homicide only relatives of the deceased could bring

the case to a court (Sealey, 1983). If the injury was a private one, then the individual himself instigated the suit (called a *dike*). In private cases only the injured (or the family of a murdered individual) could bring the charges. In the case of public harm any citizen could initiate a trial (called a *graphe*) (Glotz, 1929; MacDowell, 1978; Todd, 1993).² Furthermore, for some cases damages were either fixed by the law or determined by the judges, while in other charges the damages were not fixed (called a *timetos*) (Glotz, 1929; Hall, 1996; Todd, 1993). The Trial of Socrates was a *graphe* and a *timetos*. He had three accusers: Anytus, Lycon, and Meletus.

In *graphai* the accuser had to pay a fee for court costs. The magistrate in charge reviewed the charges, assembled the evidence furnished by the parties, and scheduled the trial (Glotz, 1929; MacDowell, 1978). They did not rule evidence or arguments out of order and had no right to direct the jury (Todd, 1993). The plaintiff could withdraw the charges at any time, but doing so resulted in a fine of 1000 drachmas and the inability to bring forth future similar cases (Glotz, 1929; MacDowell, 1978). Furthermore, private prosecutors were sanctioned if they were unable to obtain one-fifth of the juror's votes (Plato, line 36). Presumably these institutional features were designed to mitigate the prevalence of frivolous cases. MacDowell (1978) argues that this was done to deter sycophants attempting to extract bribes using the threat of prosecution. The formal charges against Socrates were “not reverencing the city's traditional gods, but new and different spiritual powers” and “corrupting the youth” (Plato, line 24; Xenophon, line 10). Since the charges included religion, initially the *archon-basileus* (king magistrate) reviewed the charges to assess whether they had merit (Stone, 1989). Furthermore, since the law did not have specific damages the charges were accompanied by a sanction recommended by the plaintiff. In Socrates' case the recommended punishment was death by drinking the poison hemlock.

In a given year 6000 citizens, chosen by lot (random draw)³ and paid,

²MacDowell (1978) discusses exceptions and provides details. For example, for certain offenses a magistrate, serving as a public official, would initiate a case; such as a general deserting his army. Also, there existed public cases, which differed procedurally from standard *graphai*.

³Bers (2000) argues that, while the typical argument for the process of selecting jurors and voting was to mitigate corruption, the symbolism and complexity was to alleviate anxiety about democratic jurors. The gods would express their wishes by guiding the luck of the draw and, thus, selection by lot was legitimate. Hence, he uses the figurative language “sortition as god”.

were installed as jurors (called dikatai) to be assigned to cases (Brickhouse and Smith, 1989; Glotz, 1929; MacDowell, 1978; Todd, 1993). Brickhouse and Smith (1989) argue that we do not know the actual size of the jury in Socrates' trial. Juries at this time typically number between a few hundred and a few thousand. It is common for authors to assume that there were 500 members of the jury (Millett, 2005; Stone, 1989). Ties were broken in favor of acquittal (MacDowell, 1978). Later Athens adopted the standard of 501-member juries.

The trial began by both the prosecution and the defense presenting their arguments and evidence. Once that was complete the jury conducts an initial vote. Each of the jurors votes to either convict or acquit the defendant knowing the recommended sanction proposed by the plaintiff. The jurors voted in secret without deliberation. If a simple majority of the jurors voted to convict, then the trial continued to a second phase. In the second phase the defendant made a counterproposal offering a second sanction for the jury to consider. Arguments were made in favor of the two proposed sanctions. Then a second vote was conducted. Again, whichever proposal received a simple majority of the vote was the one that was implemented (Brickhouse and Smith, 1989; Glotz, 1929; Stone, 1989). Each trial was completed within a day (MacDowell, 2000).

In Socrates' trial the prosecutors proposed a punishment of death to be achieved by the drinking of hemlock. As stated, it is assumed that there was 500 jurors. We do know that in the first vote if thirty jurors would have switched their vote, then Socrates would have been acquitted (Plato, line 36). Thus, it is likely that the initial vote was 280 to convict and 220 to acquit (Burnyeat, 1997). In the second phase Socrates initially proposed not only that he not be punished for his actions, but that he should be allowed to dine each day at the expense of the public at the Prytaneum, the state dining hall (Plato, line 36), which is an honor normally reserved for Olympic victors, foreign ambassadors, and served as a sacred building to honor the goddess Hestia (Stone, 1989). On the urging of his followers Socrates revised his counterproposal to allow them to pay a fine of thirty mina (Plato, line 38).⁴ Eighty more votes were cast for the prosecutor's proposal of death than

⁴In Xenophon's account he states that "when ordered to propose a counterpenalty he [Socrates] refused and forbade his friends to propose one" (Xenophon, line 23). In Plato's account Socrates does not wish to propose a penalty, but revises his proposal to allow his followers, Plato, Crito, Critobolus, and Apollodorus, to pay the fine. Socrates himself would not pay anything.

the original vote for conviction (MacDowell, 1978; Stone, 1989), which would imply 360 votes for death and 140 for Socrates' proposal.⁵ Thus, Socrates received a penalty of death.

3 Model

Consider a game with $N+2$ players. There are N jurors labeled $i = 1, 2, \dots, N$ and, for convenience, assume N is odd. Additionally, there is a prosecutor, P , and a defendant, D . Over the course of the trial evidence is presented by both P and D to the jury. This includes both evidence as to the seriousness of the harm caused as well as the likelihood of guilt. Consider, first, the harm caused. Let $H = [0, \bar{h}]$ be the set of possible levels of harm caused by the act. An act $h = 0$ represents no harm caused and $h = \bar{h}$ represents an upper bound - the cruelest act possible. Given the evidence presented during the trial Juror i assesses the likelihood that a level of harm h has been inflicted as $g_i(h)$. The case where the level of harm is perfectly observable to Juror i is simply the case where $g_i(h') = 1$ and $g_i(h) = 0 \forall h \neq h'$ where h' is the true level of harm. Second, each juror assesses the likelihood of guilt. Let the probability that Juror i believes that D committed the act that resulted in a level of harm h , conditional on the actual level of harm being h , be $\tilde{g}_i(h)$. The special case of a juror knowing with certainty that D is innocent of causing any amount of harm is $\tilde{g}_i(0) = 1$ and $\tilde{g}_i(h) = 0 \forall h \neq 0$. As a consequence, the probability that D caused an amount of harm h , assessed by Juror i , is $f_i(h) = g_i(h) \tilde{g}_i(h)$. Consequently, i 's assessment of the expected harm caused by D is

$$s_i = \int_0^{\bar{h}} h f_i(h) dh. \quad (1)$$

Denote the outcome of the trial as the sanction s . Each juror considers the evidence presented at trial and calculates an expected harm, s_i . Each juror is assumed to have single-peaked preferences where the closer the outcome of the trial is to its expected harm, the greater its payoff. Specifically, assume $u_i(s) = -|s - s_i|$. One may think of a juror's objective is to have the

⁵This fact should be viewed with skepticism. Brickhouse and Smith (1989) argue that the only evidence of this change in the vote is from Diogenes Laertius; whose account of the trial differs significantly from Plato's account. They argue that if indeed Socrates revised his counterproposal to a fine it seems unreasonable that jurors who initially voted to acquit would switch their vote to death (p.230-232).

punishment fit the crime. This takes into consideration both what crime was committed as well as the likelihood that the defendant committed that crime. The closer the sanction is to the expected harm the more satisfied a juror is. D 's objective is to minimize the punishment, $u_D(s) = -s$. P 's objective is to maximize the punishment, $u_P(s) = s$.⁶

Consider the following sequence of actions. First, P announces a sanction $s_P \in H$. Second, each juror hears the evidence, calculates s_i , and casts a vote of either 0 or 1. Let v_i^1 denote the vote cast by Juror i . If $\sum_{i=1}^N v_i^1 < \frac{N+1}{2}$, which is an integer since N is odd, then D is acquitted and $s = 0$. If $\sum_{i=1}^N v_i^1 \geq \frac{N+1}{2}$, then D announces a sanction $s_D \in H$. Fourth, each juror casts a second vote of either 0 or 1. Let v_i^2 denote the second vote cast by Juror i . If $\sum_{i=1}^N v_i^2 < \frac{N+1}{2}$, then D 's proposed punishment is accepted and $s = s_D$. If $\sum_{i=1}^N v_i^2 \geq \frac{N+1}{2}$, then P 's proposed punishment is accepted and $s = s_P$.

4 Analysis

Consider the subgame perfect Nash equilibrium of the game where no player selects a weakly dominated strategy.

4.1 Equilibrium

First, consider the second vote by the jurors. As a consequence of the requirement that no player selects a weakly dominated strategy, Juror i selects $v_i^2 = 1$ if $|s_P - s_i| < |s_D - s_i|$ and $v_i^2 = 0$ if $|s_P - s_i| \geq |s_D - s_i|$.⁷ An important reference point is the median-juror's assessment of the harm caused by D , or rather, Juror m where $\#\{i|s_i < s_m\} = \#\{i|s_i > s_m\} = \frac{N-1}{2}$. It follows immediately that whichever proposal is closer to s_m is the one that wins the vote. If s_P is greater than s_m , then D chooses a sanction less than and equidistant from s_m . This is the lowest proposed sanction that wins the jury's second vote. If s_P is less than s_m , then any proposal farther away from s_m is a best response, since all lose the vote and any closer would result in a more severe

⁶The assumption of single-peaked preferences for the jurors is important for the analysis. A relaxation of the defendant's objective function is considered in Subsection 4.3.

⁷Assume that if indifferent a juror sides with the defendant and selects $v_i^t = 0 \forall t \in \{1, 2\}$. This is, in fact, the tie-breaking rule used by the Athenians (MacDowell, 1978) when there was an equal number of votes cast for the two alternatives.

sanction. Denote, $I = \left[\max \{0, s_m - |s_P - s_m|\}, \min \{\bar{h}, s_m + |s_P - s_m|\} \right]$ as the interval of sanctions that are closer to s_m than s_P and, thus, proposals that D can make and win.

Lemma 1 *If $s_P > s_m$, then $s_D^* = \max \{0, s_m - |s_P - s_m|\}$. If $s_P \leq s_m$, then $s_D^* \notin I$.*

Proof. Consider a ultimate subgame with s_P and s_D proposed by P and D respectively (with a conviction in the initial vote). Let $v_{-i}^2 = (v_1^2, \dots, v_{i-1}^2, v_{i+1}^2, v_N^2)$ denote the profile of votes cast by all jurors other than Juror i in this final subgame. Suppose $\sum_{j \neq i} v_j^2 \neq \frac{N-1}{2}$ so that i 's vote does not have the potential to influence the outcome. Consequently, $v_i^2 = 1$ and $v_i^2 = 0$ result in the same payoff to i . Instead, suppose $\sum_{j \neq i} v_j^2 = \frac{N-1}{2}$ so that $v_i^2 = 1$ results in a payoff of $-|s_P - s_i|$ whereas $v_i^2 = 0$ generates $-|s_D - s_i|$. Thus, if $|s_P - s_i| < |s_D - s_i|$ then $v_i^2 = 0$ is weakly dominated, while if $|s_P - s_i| > |s_D^* - s_i|$ then $v_i^2 = 1$ is weakly dominated. Hence, for any v_{-i}^2 , if $|s_P - s_i| < |s_D - s_i|$ then $v_i^2 = 1$ is the equilibrium action, while if $|s_P - s_i| > |s_D^* - s_i|$ then $v_i^2 = 0$ is.

Now consider a penultimate subgame with s_P proposed by P (and a conviction in the initial vote). Suppose, first, that $s_P > s_m$. If $s_D \in I$, then, by construction, $|s_P - s_m| \geq |s_D - s_m|$ and Juror m selects $v_m^2 = 0$. Furthermore, since $\#\{i | s_i < s_m\} = \#\{i | s_i > s_m\} = \frac{N-1}{2}$, $\sum_i v_i^2 < \frac{N+1}{2}$ and the implemented sanction is s_D . If $s_D \notin I$, then, following the same argument, Juror m selects $v_m^2 = 1$ and $\sum_i v_i^2 \geq \frac{N+1}{2}$ resulting in the sanction s_P being imposed. As a result, D selects the minimal element of I , $s_D^* = \max \{0, s_m - |s_P - s_m|\}$. Instead, suppose that $s_P \leq s_m$. Every sanction in I results in a lower payoff for D than s_P and every $s_D \notin I$ results in s_P being the implemented sanction. Hence, every $s_D^* \notin I$ is a best response. ■

Now consider the first vote by the jurors. Suppose that s_P is greater than s_m . Each juror knows that if they vote to convict, then s_D^* will be the resulting sanction. Thus, a juror votes to convict if s_D^* is closer to its ideal sanction than 0. Similarly, if s_P is less than s_m , then it knows that s_P will be the result of a conviction in the initial vote. Thus, a juror votes to convict if s_P is closer than 0.

P , on the other hand, selects a proposal that results in the highest sanction possible being implemented. If s_P is chosen to be greater than s_m , then s_D^* wins the second vote. Since s_D^* is equidistant from s_m , a decrease in s_P increases the resulting sanction imposed on D . Additionally, if s_P is chosen to

be less than s_m , then s_P is the outcome of the second vote and an increase in s_P increases the imposed punishment. Consequently, the best sanction is the median-juror's ideal sanction and a unique subgame perfect Nash equilibrium outcome where no individual selects weakly dominated strategies arises.

Proposition 1 *In the unique subgame perfect Nash equilibrium outcome where no individual selects a weakly dominated strategy the prosecutor proposes the median-juror's most preferred sanction, $s_P^* = s_m$, the jury votes to convict, and then imposes the median-juror's ideal sanction.*

Proof. Given the behavior described in Lemma 1 for the ultimate and penultimate subgames, consider the initial vote. Consider Juror i 's decision given v_{-i}^1 . If $\sum_{j \neq i} v_j^1 = \frac{N-1}{2}$, then $v_i^1 = 1$ results in a payoff of $-|s_P - s_i|$ if $s_P \leq s_m$ and $-|\max\{0, s_m - |s_P - s_m|\} - s_i|$ if $s_P > s_m$, while $v_i^1 = 0$ generates $-|0 - s_i|$. If $\sum_{j \neq i} v_j^1 \neq \frac{N-1}{2}$, then i 's payoff is the same for v_i^1 of 0 or 1. Hence, if $|s_P - s_i| < |0 - s_i|$ when $s_P \leq s_m$ or $|\max\{0, s_m - |s_P - s_m|\} - s_i| < |0 - s_i|$ when $s_P > s_m$, then $v_i^1 = 1$ is Juror i 's equilibrium action. Finally, consider the selection of $s_P^* = s_m$. At this proposal $v_m^1 = 1$ and since $\#\{i | s_i < s_m\} = \#\{i | s_i > s_m\} = \frac{N-1}{2}$, $\sum_i v_i^1 > \frac{N-1}{2}$. The proposal $s_P^* = s_m$ generates a greater payoff than all $s_P > s_m$ since $\max\{0, s_m - |s_P - s_m|\}$ is maximized at $s_P = s_m$. It also generates a greater payoff than any $s_P < s_m$ since any such proposal would win, but be less. Thus, $s_P^* = s_m$ is the unique best response. Hence, from Lemma 1, $\sum_i v_i^1 > \frac{N-1}{2}$ and $s_D^* \notin I$, so that $\sum_i v_i^1 > \frac{N-1}{2}$, describes the equilibrium path. ■

This median-juror result is simply an extension of the median-voter theorem (Black, 1948) to jury decisionmaking. The single-peaked preference assumption is natural in the context of jurors attempting to have the punishment fit the crime. An appealing property of the median juror's ideal sanction is that it is a Condorcet Winner, that is, paired against any other sanction a majority of the juror's prefer it. Thus, one may argue that if the jurors, by the virtue of their being exposed to the evidence in the trial, are the best informed individuals, then the structure of the trials in 399 B.C. Athens produce a desirable outcome.

Furthermore, it has been argued that sanctions imposed should depend on the expected harm and the certainty of guilt (Andreoni, 1991; Lando, 2005; Miceli, 1990; Rubinfeld and Sappington, 1987). First, consider two trials where $g_i^1(h) = g_i^2(h) \forall i \forall h$ so that the harm caused is identical, but

the jurors are more certain of the defendant's guilt in the first. Formally, suppose that the evidence in the first trial first-order stochastically dominates the second trial. Thus, $s_i^1 > s_i^2 \forall i$ and, consequently, $s_m^1 > s_m^2$. Similarly, if $\tilde{g}_i^1(h) = \tilde{g}_i^2(h) \forall i \forall h$, but the evidence of harm in the first trial first-order stochastically dominates the second trial, then, again, $s_m^1 > s_m^2$. Thus, the Trial of Socrates hands out stronger sanctions for individuals more likely to be guilty as well as those who caused a higher expected harm.

Finally, note that the maximal sanction, \bar{h} , occurs if and only if the median juror believes with certainty that the harm caused was maximal, $g_m(\bar{h}) = 1$, and believes with certainty that the defendant is the one who caused this maximal harm, $\tilde{g}_m(\bar{h}) = 1$.

4.2 Expected Sanction and Deterrence

It is also important to understand the impact on an individual's incentive to break the law. From the standard model of optimal enforcement and deterrence, each individual weighs the benefit from committing an illegal act with the expected sanction imposed (Becker, 1968; Garoupa, 1997; Polinsky and Shavell, 2000). If the expected sanction is greater, then the overall amount of criminal activity in the population is less.

If an individual D engages in an activity, he assesses that a harm h is inflicted with probability $g_D(h)$. Also, with probability $q(h)$ he is apprehended and stands trial. The probability of apprehension may depend on the level of harm caused. For example, minor acts with little impact on others may not be devoted the enforcement resources. Let $f_D(h) = q(h)g_D(h)$ denote the probability that D inflicts a harm h and is apprehended.⁸ Therefore, the expected sanction if choosing to break the law is

$$E_{TOS} = s_m \int_0^{\bar{h}} f_D(h) dh. \quad (2)$$

The Trial of Socrates differs in two important ways. First, a simple majority vote is used rather than a unanimous vote. Second, the jury decides whether to implement the prosecutor's or the defendant's proposed sanction, rather

⁸While $\int g_i(h) dh = 1 \forall i = 1, \dots, N, D$, it need not be the case that $\int q(h) dh = 1$. Rather, the probability of apprehension can be less than one. Consequently, $\int f_D(h) dh \in (0, 1)$. If the probability of apprehension is independent of the harm inflicted, $q(h) = \bar{q} \forall h$, then $\int f_D(h) dh = \bar{q}$.

than deciding between acquittal and a fixed sanction. Consider, then, a unanimous vote required to implement a sanction \bar{s} , rather than acquit. One may think of a punishment for a crime fixed by statute (as in habitual offender laws) or the expected sanction chosen by the judge (which may, for example, center around Federal sentencing guidelines). The sanction \bar{s} is implemented if and only if every juror prefers \bar{s} to 0. Let s_{min} be defined such that $s_{min} = \min \{s | s \in S\}$ where $S = \{s_1, s_2, \dots, s_N\}$. Friedman and Wickelgren (2006) studying the effect of Bayesian jurors on crime consider a similar setup where the jury set threshold for which the probability of guilt must exceed in order to convict. With a heterogenous jury the standard is the jury least likely to vote to convict. It follows that a conviction with unanimous voting requires that $\bar{s} \leq 2s_{min}$ so that even the juror with the lowest assessed s_i prefers \bar{s} to 0. Any fixed sanction greater than this results in an acquittal. Furthermore, to a potential criminal the expected sanction is

$$E_U = \begin{cases} \bar{s} \int_0^{\bar{h}} f_D(h) dh & \text{if } \bar{s} \leq 2s_{min} \\ 0 & \text{if } \bar{s} > 2s_{min} \end{cases} \quad (3)$$

Hence, $E_U < E_{TOS}$ if either the fixed sanction is too severe for the “minimum juror”, $\bar{s} > 2s_{min}$, or the fixed sanction is not too severe and, in fact, is less severe than the median juror’s ideal sanction, $s_m > \bar{s}$. Thus, if the minimum juror prefers an acquittal to the fixed sanction or the fixed sanction is less than the median juror’s ideal sanction, then the Trial of Socrates results in a more deterrence. Only if the fixed sanction is greater than the median juror’s ideal sanction, but less than twice the size of the minimum juror’s ideal sanction would the Trial of Socrates deter less.

4.3 Socrates’ Death

An immediate potential critique of the previous work is that the model predicts that if and only if both $g_m(\bar{h}) = 1$ and $\tilde{g}_m(\bar{h}) = 1$ would a maximal sanction be the outcome. The one data point we have of this type of trial being used is Socrates’ trial. In this trial the maximal sanction was imposed. How might this outcome be rationalized?

An assumption used in the formal model is that the objective function of the defendant is decreasing in the implemented sanction, $u_D(s) = -s$. There is a significant amount of evidence and argument that this payoff function, while likely an accurate depiction of most defendants, does not describe Socrates.

Pangle (1985) argues that Socrates’s goal, after the initial conviction, was to be sentenced to death. The author is interested in the means by which Socrates achieved this objective. Any number of actions could have been selected that all resulted in death. Pangle (1985) provides evidence that Socrates’ antagonizing speech was designed to argue that he was not impious and unjust and still be sentenced to die. Danzig (2003) argues that Socrates, at least the Socrates depicted by Xenophon, did in fact not fail at the trial, but achieved his objective of death. Danzig (2003) argues that a proper interpretation of the accounts of the trial, Plato’s *Apology* and Xenophon’s *Apology*, are best thought of as responses to after-trial criticisms that Socrates failed at his defense.

In Xenophon’s account of the trial Socrates specifically states that if “I had succeeded [at the defense], it’s clear that I would have managed to exchange a quick death for a painful one” (line 8) and his strategy was to “offend the jury and choose death like a free man rather than slavishly beg for the worthless gain of continued life” (line 9). Xenophon himself states that “Socrates thought death was now preferable to life” (line 1) and “believed that the time had come for him to die” (line 23). Even in Plato’s account Socrates claims that “death is good” and that “I count it [death] as a gain” (line 40). Therefore, it seems that while it is natural to assume that most defendants would prefer to minimize the sanction, Socrates prefers a maximal punishment.⁹

Therefore, suppose that $u_D(s) = s$. Given that P has proposed a high sanction, D does not want to make a proposal that wins and will select $s_D \notin I$. Hence, a proposal by D that is “too low”, as in the case of Socrates, is a best response. In the initial vote, each juror, selects between acquittal, generating $-|s_i - 0|$, and P ’s proposed sanction, resulting in $-|s_i - s_P|$. Therefore, the prosecutor selects the highest sanction that still results in the median juror preferring it to acquittal.¹⁰

Proposition 2 *Suppose the defendant prefers a severe sanction, $u_D(s) = s$.*

⁹Of course, if Socrates is actually guilty, as argued by Burnyeat (1997) and the harm caused is high enough to justify the sentence, then it may be that the outcome of Socrates’ trial does in fact fit with the model’s prediction.

¹⁰While not directly analyzed here, it is straightforward to see that if the defendant has single-peaked preferences his ideal sanction, if greater than the median-juror’s ideal sanction, is the outcome of the trial. Additionally, if the prosecutor has single-peaked preferences then her ideal sanction is the outcome so long as it is less than the median juror’s.

If $2s_m > \bar{h}$, then the prosecutor proposes \bar{h} , the defendant proposes a low sanction, $s_P \in [0, s_m - |s_m - \bar{h}|)$, and the jury implements a sanction of \bar{h} . Otherwise, if $2s_m \leq \bar{h}$, then the prosecutor's proposal of $2s_m$ is implemented.

Proof. Voting behavior in the ultimate subgames remains unchanged from Lemma 1. Therefore, consider the selection of s_D . As shown in the proof of Lemma 1, any $s_D \in I$ results in $s = s_D$ being implemented. A selection of $s_D \notin I$ results in the outcome $s = s_P$. Since $u_D(s) = s$, D prefers the greater of the two. Hence, if $s_P < s_m$ then $s_D^* = \min\{\bar{h}, s_m + |s_P - s_m|\}$ is the best response, while if $s_P \geq s_m$ then $s_D^* \notin I$ are best responses.

Now consider the initial vote by the jurors. Using backward induction, given any v_{-i}^1 , if $\sum_{j \neq i} v_j^1 = \frac{N-1}{2}$ then $v_i^1 = 1$ generates $-|s_P - s_i|$ if $s_P \geq s_m$ and $-\left|\min\{\bar{h}, s_m + |s_P - s_m|\} - s_i\right|$ if $s_P < s_m$, while $v_i^1 = 0$ results in $-|0 - s_i|$. If $\sum_{j \neq i} v_j^1 \neq \frac{N-1}{2}$ then i 's payoff is the same regardless of v_i^1 . Therefore, the equilibrium action for Juror i is $v_i^1 = 1$ if and only if $|s_P - s_i| < |0 - s_i|$ when $s_P \geq s_m$ and $\left|\min\{\bar{h}, s_m + |s_P - s_m|\} - s_i\right| < |0 - s_i|$ when $s_P < s_m$.

Finally, consider the initial move. If P selects $s_P \geq s_m$ then s_P is implemented if $s_P < 2s_m$; otherwise D is acquitted. If $s_P < s_m$ then $\min\{\bar{h}, s_m + |s_P - s_m|\}$ is the outcome. Therefore, if $\bar{h} \leq 2s_m$ a selection of $s_P^* = \bar{h}$ (weakly) dominates all other selections. If $\bar{h} > 2s_m$ then $s_P = 2s_m$ is the best choice when $s_P > s_m$, while $\min\{\bar{h}, s_m + |s_P - s_m|\} = 2s_m$, or rather, $s_P = 0$ (or rather $s_P = \varepsilon$ for a small ε) is the best choice when $s_P < s_m$ (so that $s_D^* = 2s_m$). P prefers the first.

Consequently, if $\bar{h} \leq 2s_m$ then an equilibrium path is $s_P^* = \bar{h}$, $I = [\max\{0, 2s_m - \bar{h}\}, \bar{h}]$, and $s_D^* < 2s_m - \bar{h}$ (or $s_D^* = \bar{h}$). Instead, if $\bar{h} > 2s_m$ then the equilibrium path is $s_P^* = 2s_m$, $I = [0, 2s_m]$, and $s_D^* > 2s_m$. ■

Consequently, if death is assumed to be the maximal sanction, then the outcome of Socrates' trial is explained by Socrates' preferring death and the median juror preferring death to acquittal.

Alternatively, of course, one could argue that Socrates' death was not the maximal sanction. Athenian law had other more severe forms of execution including, for example, crucifixion (Hall, 1996; Smith 2008). Athenian jurors were restricted to punishments of fines, disenfranchisement, exile, or execution (Hall, 1996). Bloch (2002) argues that hemlock causes a relatively calm and peaceful death. Brickhouse and Smith (1989) argue that Socrates in

Plato's account did not prefer death and, in fact, defended himself in a way that was consistent with his own beliefs. Plato's account does not support Socrates's attempting death and Xenophon was simply attempting to excuse Socrates' loss (Smith, 2008). Hence, a second explanation of Socrates' death is that the prosecutors did indeed propose the sanction preferred by the median juror (or at least the one that was feasible and closest to the median juror). It follows from Lemma 2 than any proposal by Socrates is an equilibrium since I is empty. Consequently, Socrates' death does not contradict Proposition 1.

5 Discussion

Concerns have been raised about the effectiveness of trials. The punishment may not fit the crime (Andreoni, 1991), the sanction may not reflect the degree of certainty of the guilt of the defendant (Lando, 2005; Miceli, 1990), and the possibility of acquittal may reduce the expected sanction and result in less deterrence (Friedman and Wickelgren, 2006; Kambhu, 1989). It is shown here that the Trial of Socrates improves upon all of these critiques. Thus, instituting similar mechanisms for the criminal justice system warrants further consideration.

There are a number of considerations that are not formally modeled here, which, if analyzed, would check the robustness of the outcome presented. The model assumes that the ideal sanction of each juror is known. This is unlikely. From the defendant and prosecutor's perspective what matters is their expectation of the median-juror's ideal sanction. Thus, one would expect an expected median-juror result, but this is left for future consideration. Also, a juror's ideal point is simply the expected harm caused by the defendant, as assessed by that individual juror. Thus, one may think of a juror's motive as to have the punishment fit the crime. The model could be applied to situations where the preferences depends on other factors (such as race of the defendant or personal experiences of a juror). The median-juror result will still hold in such an extension, but the responsiveness of the sanction to the harm caused will depend on the formation of the median juror's ideal sanction. Furthermore, there are four aspects of 399 B.C. Athenian justice not modeled. First, there was no public prosecutor. Charges were brought by private individuals. In Socrates' case, three men brought up the charges. Second, two mechanisms were used to protect against frivolous cases: the

archon first reviewed the charges before allowing them to go to trial and prosecutors were punished if they received less than $\frac{1}{5}$ of the juror's votes. It is not necessarily clear what impact these two institutional factors would have on both the outcome of the trial and the deterrence of criminal behavior. The third aspect of Athenian justice is that juries were typically large ranging from a few hundred to some thousand members (Brickhouse and Smith, 1989). While the results presented here are independent of the jury size (just the position of the median juror matters), consideration should be given to the effect of jury size on the environment. Finally, if the initial vote is for conviction, then not only does the defendant propose a sanction, but additional arguments are made. Hence, it is possible that in this second round beliefs regarding the appropriate sanction and certainty of guilt may change. Future consideration could be given to the impact of these additional institutional factors on the outcome of the trials.

6 References

Andreoni, James (1991), Reasonable Doubt and the Optimal Magnitude of Fines: Should the Penalty Fit the Crime?, *RAND Journal of Economics* 22(3), 385-395.

Becker, Gary S. (1968), Crime and Punishment: An Economic Approach, *Journal of Political Economy* 76(2), 169-217.

Bers, Victor (2000), Just Rituals. Why All the Rigmarole of Fourth-Century Athenian Lawcourts?, in Pernille Flensted-Jensen, Thomas Heine Nielsen, and Lene Rubinstein eds, *Polis and Politics: Studies in Ancient Greek History*, Copenhagen: Museum Tusulanum Press, p.553-562.

Black, Duncan (1948), On the Rationale of Group Decision-Making, *Econometrica* 56(1), 23-34.

Bloch, Enid (2002), Hemlock Poisoning and the Death of Socrates: Did Plato Tell the Truth?, in Thomas C. Brickhouse and Nicholas D. Smith eds, *The Trial and Execution of Socrates: Sources and Controversies*, New York: Oxford University Press, 255-278.

Brickhouse, Thomas C. and Nicholas D. Smith (1989), *Socrates on Trial*, Princeton: Princeton University Press.

Burnyeat, M. F. (1997), The Impiety of Socrates, *Ancient Philosophy* 17, 1-12.

Connolly, Terry (1987), Decision Theory, Reasonable Doubt, and the Utility of Erroneous Acquittals, *Law and Human Behavior* 11(2), 101-112.

Danzig, Gabriel (2003), Apologizing for Socrates: Plato and Xenophon on Socrates' Behavior in Court, *Transactions of the American Philological Association* 133(2), 281-321.

Farquharson, Robin (1969), *Theory of Voting*, New Haven: Yale University Press.

Friedman, Ezra and Abraham L. Wickelgren (2006), Bayesian Juries and the Limits to Deterrence, *Journal of Law, Economics, & Organization* 22(1), 70-86.

Garoupa, Nuno (1997), The Theory of Optimal Law Enforcement, *Journal of Economic Surveys* 11, 267-295.

Glötz, G. (1929), *The Greek City and Its Institutions*, translated by N. Mallinson, London: Kegan Paul, Trench, Trubner, and Co.

Hall, Margaretha Debrunner (1996), Even Dogs Have Erinyes: Sanctions in Athenian Practice and Thinking, in Lin Foxhall and Andrew D. E. Lewis eds, *Greek Law in its Political Setting*, Oxford: Clarendon Press, 73-89.

Kambhu, John (1989), Regulatory Standards, Noncompliance and Enforcement, *Journal of Regulatory Economics* 1, 103-114.

Lando, Henrik (2005), The Size of the Sanction Should Depend on the Weight of the Evidence, *Review of Law and Economics* 1(2), 277-292.

MacDowell, Douglas M. (1978), *The Law in Classical Athens*, Ithaca: Cornell University Press.

MacDowell, Douglas M. (2000), The Length of Trials for Public Offenses in Athens, in Pernille Flensted-Jensen, Thomas Heine Nielsen, and Lene Rubinstein eds, *Polis and Politics: Studies in Ancient Greek History*, Copenhagen: Museum Tusulanum Press, p.563-568.

- Miceli, Thomas J. (1990), Optimal Prosecution of Defendants Whose Guilt is Uncertain, *Journal of Law, Economics, & Organization* 6(1), 189-210.
- Millett, Paul (2005), The Trial of Socrates Revisited, *European Review of History* 12(1), 23-62.
- Pangle, Thomas L. (1985), The Political Defense of Socratic Philosophy: A Study of Xenophon's "Apology of Socrates to the Jury", *Polity* 18(1), 98-114.
- Plato, *Apology*, translated by Raymond Larson in *The Apology and Crito of Plato and The Apology and Symposium of Xenophon*, Lawrence, Kansas: Coronado Press.
- Polinsky, A. Mitchell and Steven Shavell (2000), The Economic Theory of Public Enforcement of Law, *Journal of Economic Literature* 38(1), 45-78.
- Riker, William (1986), *The Art of Political Manipulation*, New Haven: Yale University Press.
- Rubinfeld, Daniel L. and David E. M. Sappington (1987), Efficient Awards and Standards of Proof in Judicial Proceedings, *RAND Journal of Economics* 18(2), 308-318.
- Sealey, Raphael (1983), The Athenian Courts for Homicide, *Classical Philology* 78(4), 275-296.
- Smith, Nicholas D. (2008), personal communication, April X.
- Stone, I. F. (1989), *The Trial of Socrates*, New York: Doubleday.
- Todd, Stephen C. (1993), *The Shape of Athenian Law*, Oxford: Clarendon Press.
- Xenophon, *Apology*, translated by Raymond Larson in *The Apology and Crito of Plato and The Apology and Symposium of Xenophon*, Lawrence, Kansas: Coronado Press.