A Theory of Hung Juries and Informative Voting*

Fuhito Kojima[†] Yuki Takagi[‡]

April 28, 2009

Abstract

This paper investigates a jury decision when hung juries and retrials are possible. When jurors in subsequent trials know that previous trials resulted in hung juries, informative voting cannot be an equilibrium regardless of voting rules unless the probability that each juror receives the correct signal when the defendant is guilty is identical to the one when he is innocent. Thus, while Coughlan (2000) claims that mistrials facilitate informative voting, our result shows that such an assertion holds only in limited circumstances.

1 Introduction

In many jurisdictions in the United States and elsewhere, unanimity among jurors is required for jury verdict. The unanimity rule is commonly believed

*We are grateful to Alberto Alesina, James Alt, Attila Ambrus, Georgy Egorov, Drew Fudenberg, John William Hatfield, Shin Kanaya, Yukio Koriyama, Scott Moser, Parag Pathak, John Patty, Elizabeth Maggie Penn, Kenneth Shepsle, Satoru Takahashi, Leeat Yariv and participants in Research Workshop in Political Economy and Research Workshop in American Politics at Harvard University and MPSA 2009 for comments, and to Ben Edelman, Meghan Ferrer, Gregor Matvos, Anthony Niblett and Giacomo Ponzetto for answering our questions about the judicial system in the United States. We appreciate comments from an anonymous referee and the associate editor that greatly improved the paper.

[†]Cowles Foundation, Yale University, New Haven, CT 06510, e-mail: fuhitokojima1979@gmail.com.

[‡]Corresponding author. Department of Government, Harvard University, Cambridge, MA 02138, e-mail: takagi@fas.harvard.edu.

to minimize the possibility of convicting an innocent defendant. This view is challenged by Feddersen and Pesendorfer (1998) in light of game-theoretic analysis of voting behavior. They show that, if jurors vote strategically, the unanimity rule may convict the innocent and acquit the guilty more often than most rules, including the simple majority rule.

Coughlan (2000) provides a counterargument to their claim in his study of jury decision with mistrials. He points out that unanimity is required for either conviction or acquittal in many jurisdictions. Otherwise a hung jury occurs and the case faces a new trial in the future with a new group of jurors.[1] Based on this idea, he constructs a model in which informative voting is an equilibrium for a nontrivial range of parameters.

The above model, however, does not explicitly allow information transmission between trials. By contrast, information is often disclosed between trials in reality. For example, news media regularly covers trials and information about mistrials is frequently reported.

To capture this feature, this paper allows jurors to know that previous trials (if any) resulted in hung juries and use this fact to infer the likelihood of guilt or innocence of the defendant. For any voting rule, we show that informative voting is an equilibrium only in knife-edge cases where the probability that a juror receives the correct signal when the defendant is guilty is exactly the same as the one when the defendant is innocent. Although Coughlan argues that mistrials facilitate informative voting, our result demonstrates that his claim has only limited applicability.

In addition to mistrials, Coughlan (2000) presents another model to justfy unanimity rule. He introduces a model with deliberation and shows that informative voting is an equilibrium for a nontrivial range of parameters. Austen-Smith and Feddersen (2006) further assume uncertainty about juror preferences. They show that informative voting is not an equilibrium under the unanimity rule if jurors have heterogeneous preferences, while it may be an equilibrium under nonunanimous rules. Both their paper and ours question the efficacy of the unanimity rule in realistic models: deliberation in their paper and hung juries in ours.

The current work is part of the literature on jury design under strategic voting pioneered by Austen-Smith and Banks (1996). They study the

[1] A mistrial may be declared for a number of other reasons, such as juror misconduct. However we focus on a mistrial that occurs because of a hung jury and use these terms interchangeably.

strategic aspect of jury decision and show that informative voting often fails to be an equilibrium. McLennan (1998) and Wit (1998) show that there exists an equilibrium, which may not be informative, that aggregates jurors' information in an adequate manner. Duggan and Martinelli (2001) and Meirowitz (2002) consider continuous signals. Guarnaschelli, McKelvey, and Palfrey (2000) find experimental evidence of strategic voting in jury setting. Study of Condorcet Jury theorems has a long tradition in a more statistically oriented literature (Grofman and Feld, 1988; Young, 1988; Ladha, 1992).

2 Model and Result

A defendant is under a jury trial on a criminal charge. There are two states of the world $\omega \in \{G, I\}$: the defendant is either guilty (denoted G) or innocent (I). The prior probabilities that the state is G and I are $r \in (0,1)$ and $1-r$, respectively.

At each period $t = 0, 1, 2, \ldots$, a jury composed of n jurors is formed. The jury at t, denoted by N^t,[2] makes a decision $d_t \in \{A, C, M\}$ following a voting rule specified by an integer $\hat{k} \in (n/2, n]$. At each period t, each juror $j \in N^t$ casts a vote $v_j \in \{a, c\}$, where a is a vote for acquittal and c is a vote for conviction. If at least $\hat{k}$ jurors vote for conviction, then the defendant is convicted ($d_t = C$); if at least $\hat{k}$ jurors vote for acquittal, then the defendant is acquitted ($d_t = A$); if neither of these happens, then a mistrial is declared ($d_t = M$). When a mistrial is declared, a new jury (with new members) is formed at period $t+1$ and the same procedure is repeated. Once the jury reaches a verdict ($d_t \in \{A, C\}$) at any t, it becomes the final decision $d_\infty = d_t \in \{A, C\}$.

Each juror j receives private signal $s_j \in \{g, i\}$ correlated with the state of the world $\omega \in \{G, I\}$. Given ω, s_j follows an i.i.d. distribution. Let p_g (respectively p_i) be the conditional probability that each juror observes the "correct signal" g (respectively i) when the true state is G (respectively I). We assume $p_g, p_i \in (1/2, 1)$.

The utility of juror j when the state is ω and a final decision $d_\infty \in \{C, A\}$ is made is denoted by $u_j(d_\infty, \omega)$, and defined as $u_j(C, G) = u_j(A, I) = 0$, $u_j(C, I) = -q_j$, $u_j(A, G) = -(1-q_j)$, where $q_j \in (0,1)$. When an infinite

[2]For example, the first jury is denoted by N^0, the jury after one mistrial by N^1, and so on.

sequence of mistrials occurs, each juror receives some fixed utility.[3] Juror j prefers conviction to acquittal if and only if she places at least probability q_j that the defendant is guilty. We assume that $\bar{q} \equiv \sup_{j \in \bigcup_{t=0}^{\infty} N^t} q_j < 1$ and $\underline{q} \equiv \inf_{j \in \bigcup_{t=0}^{\infty} N^t} q_j > 0$. In words, while we allow for an infinite number of potential jurors, levels of "reasonable doubt" for the population is bounded away from the extreme values, 0 and 1.[4]

All jurors in the t^{th} trial know that there were $t-1$ trials before and all of them resulted in mistrials, but they do not know any further information from previous trials. We assume that there is at least one numerical split under which a mistrial is declared. This assumption holds except for simple majority rule with n odd.

A strategy of each juror j is a function which assigns a vote v_j for each private signal s_j. We say that a strategy profile is an **informative voting** if each juror j chooses c if $s_j = g$ and a if $s_j = i$.

When the signals for guilt and innocence are equally accurate, that is, $p_g = p_i = p$ for some $p \in (1/2, 1)$, a straightforward adaptation of Coughlan (2000) shows that informative voting forms a sequential equilibrium if and only if, for each juror j,

$$\frac{r}{1-r} \cdot \frac{1-p}{p} \leq \frac{q_j}{1-q_j} \leq \frac{r}{1-r} \cdot \frac{p}{1-p}.$$

Next, consider the case with $p_g \neq p_i$.

Theorem 1. *Suppose $p_g \neq p_i$. Then informative voting never forms a sequential equilibrium.*

[3]Since our analysis focuses on informative voting, infinite mistrials occur with probability zero. Hence no specific assumption on the utility for infinite mistrials is important for the analysis. Also we note that retrials may be costly and prosecutors may decide not to seek a retrial after a mistrial with some probability in reality. Consider a mode in which the game proceeds as in the paper except that there is $\mu \in (0, 1)$ such that, every time a hung jury occurs, the defendant is acquitted with probability μ without facing a retrial. Theorem 1 carries over to such a model since each juror's posterior is unaffected by this change in the game.

[4]This assumption is satisfied quite broadly. One example is a model in which the parameter q_j is chosen from a finite set of possible values in $(0, 1)$. A stationary case in which the utility characteristic of each jury is the same for every period is a particular case of such a specification. Even if the population of potential jurors does not satisfy the bound, jury selection procedures (often called voir dire) may render the bound plausible, since the processes prevent highly biased individuals from serving in juries.

We defer the proof to the Appendix and offer an intuition here. Suppose, for instance, that the signal of guilt is more accurate than that of innocence ($p_g > p_i$). Then the jurors are more likely to receive mistaken signals and, as a result, fail to agree on a verdict when the defendant is innocent than when he is guilty (Lemma 1 in the Appendix). Hence, if a juror knows that there was a hung jury before the current trial, she infers that the defendant is more likely to be innocent by Bayes' law. If hung juries occur at sufficiently many periods, then information from having these previous hung juries will become so strong that a juror is willing to vote for acquittal even if she observes a guilty signal. Thus informative voting fails to be an equilibrium.

Theorem 1 suggests that informative voting is rarely an equilibrium. Note that $p_g \neq p_i$ holds generically. Moreover, the Theorem applies to any voting rule which allows for a hung jury (i.e., the only exception is simple majority rule with n odd). Thus the defense of unanimity rule by Coughlan is inapplicable to most cases.

The conclusion of Theorem 1 can be strengthened if additional information is available to jurors. For example, it can be shown that informative voting is not an equilibrium regardless of probabilities of the correct signal if jurors are informed of numerical splits of votes in previous trials.[5]

Appendix: Proof of Theorem 1

For any $p \in [0, 1]$, let

$$\pi_M(p) = \sum_{k=n-\hat{k}+1}^{\hat{k}-1} \binom{n}{k} p^k (1-p)^{n-k},$$

$$\pi_C(p) = \sum_{k=\hat{k}}^{n} \binom{n}{k} p^k (1-p)^{n-k},$$

and $\pi_A(p) = 1 - \pi_M(p) - \pi_C(p)$. For $d \in \{M, C, A\}$, $\pi_d(p)$ is the probability that a single jury composed of n jurors makes a decision d if each juror independently votes for conviction with probability p. It is clear by definition that $\pi_M(p) = \pi_M(1-p)$ for any p. The following lemma is key for our analysis.

Lemma 1. *$\pi_M(p)$ is strictly decreasing in $p \in (1/2, 1]$.*

[5]The proof is omitted, but is available from the authors upon request.

Proof. By differentiation,

$$\begin{aligned}
&\frac{d\pi_M(p)}{dp} \\
&= \sum_{k=n-\hat{k}+1}^{\hat{k}-1} \binom{n}{k} k p^{k-1}(1-p)^{n-k} - \sum_{k=n-\hat{k}+1}^{\hat{k}-1} \binom{n}{k}(n-k)p^k(1-p)^{n-k-1} \\
&= \sum_{k=n-\hat{k}}^{\hat{k}-2} \binom{n}{k+1}(k+1)p^k(1-p)^{n-k-1} - \sum_{k=n-\hat{k}+1}^{\hat{k}-1} \binom{n}{k}(n-k)p^k(1-p)^{n-k-1} \\
&= \sum_{k=n-\hat{k}+1}^{\hat{k}-2} \left(\binom{n}{k+1}(k+1) - \binom{n}{k}(n-k)\right) p^k(1-p)^{n-k-1} \\
&\quad + \binom{n}{n-\hat{k}+1}(n-\hat{k}+1)p^{n-\hat{k}}(1-p)^{\hat{k}-1} \\
&\quad - \binom{n}{\hat{k}-1}(n-\hat{k}+1)p^{\hat{k}-1}(1-p)^{n-\hat{k}}. \qquad (1)
\end{aligned}$$

By a well-known identity $\binom{n}{k+1}(k+1) = \binom{n}{k}(n-k)$, the terms in summation in equation (1) cancel each other out. By this and the fact $\binom{n}{n-\hat{k}+1} = \binom{n}{\hat{k}-1}$,

$$\begin{aligned}
&\frac{d\pi_M(p)}{dp} \\
&= \binom{n}{\hat{k}-1}(n-\hat{k}+1)p^{n-\hat{k}}(1-p)^{\hat{k}-1} - \binom{n}{\hat{k}-1}(n-\hat{k}+1)p^{\hat{k}-1}(1-p)^{n-\hat{k}} \\
&= \binom{n}{\hat{k}-1}(n-\hat{k}+1)[p^{n-\hat{k}}(1-p)^{\hat{k}-1} - p^{\hat{k}-1}(1-p)^{n-\hat{k}}]. \qquad (2)
\end{aligned}$$

The assumption that there is at least one numerical split under which a mistrial is declared implies $\hat{k} > (n+1)/2$ and hence $\hat{k}-1 > n-\hat{k}$. Also, the assumption $p \in (1/2, 1]$ implies $p > 1-p$. Thus expression (2) is always negative. Therefore $\pi_M(p)$ is strictly decreasing in $p \in (1/2, 1]$. □

Proof of Theorem 1. First consider the case with $p_g > p_i$. Let m_j^ω be the expected disutility of juror j when the jury which contains j is hung, while each juror follows an informative voting and the true state is ω.[6]

[6] m_j^ω can be calculated by $m_j^G = (1-q_j)\pi_A(p_g) + m_j^G \pi_M(p_g)$ and $m_j^I = q_j \pi_C(1-p_i) + m_j^I \pi_M(1-p_i)$.

For an arbitrary j, let $\Pr(\omega = \bar{\omega}, |s|_{n-1} = \hat{k} - 1 | s_j, t)$ be the probability that the state is $\bar{\omega}$ and, of $n-1$ jurors other than j, $\hat{k}-1$ jurors observe signal s, conditional on the event that juror $j \in N^t$ observes signal s_j and the existence of all the previous mistrials. Since jurors in N^t observe t hung juries at periods $0, \ldots, t-1$, we obtain, for example,

$$\Pr(\omega = G, |g|_{n-1} = \hat{k} - 1 | g, t) = \frac{r\binom{n-1}{\hat{k}-1}[\pi_M(p_g)]^t p_g^{\hat{k}}(1-p_g)^{n-\hat{k}}}{r[\pi_M(p_g)]^t p_g + (1-r)[\pi_M(1-p_i)]^t(1-p_i)}.$$

Define $EU_j(v_j|s_j)$ to be the expected utility for juror j of casting a vote v_j, conditional on j getting to vote and observing signal s_j. Juror $j \in N^t$ who observes $s_j = g$ has incentives to vote for conviction if and only if $EU_j(c|g) - EU_j(a|g) \geq 0$. Note that

$$\begin{aligned}
EU_j(c|g) - EU_j(a|g) &= [-q_j - (-m_j^I)]\Pr(\omega = I, |g|_{n-1} = \hat{k}-1|g,t) \\
&\quad + [-m_j^G - (-(1-q_j))]\Pr(\omega = G, |i|_{n-1} = \hat{k}-1|g,t) \\
&\quad + [-m_j^I - 0]\Pr(\omega = I, |i|_{n-1} = \hat{k}-1|g,t) \\
&\quad + [0 - (-m_j^G)]\Pr(\omega = G, |g|_{n-1} = \hat{k}-1|g,t) \\
&\leq [-\underline{q} - (-m_j^I)]\Pr(\omega = I, |g|_{n-1} = \hat{k}-1|g,t) \\
&\quad - [m_j^G - (1-q_j)]\Pr(\omega = G, |i|_{n-1} = \hat{k}-1|g,t) \\
&\quad - m_j^I \Pr(\omega = I, |i|_{n-1} = \hat{k}-1|g,t) \\
&\quad + m_j^G \Pr(\omega = G, |g|_{n-1} = \hat{k}-1|g,t).
\end{aligned}$$

By Lemma 1, $p_g > p_i$ implies $\pi_M(p_g) < \pi_M(p_i) = \pi_M(1-p_i)$. Therefore $\lim_{t\to\infty}\Pr(\omega = G, |g|_{n-1} = \hat{k}-1|g,t) = \lim_{t\to\infty}\Pr(\omega = G, |i|_{n-1} = \hat{k}-1|g,t) = 0$. Moreover $\lim_{t\to\infty}\Pr(\omega = I, |g|_{n-1} = \hat{k}-1|g,t) = \binom{n-1}{\hat{k}-1}(1-p_i)^{\hat{k}-1}p_i^{n-\hat{k}} < \binom{n-1}{\hat{k}-1}p_i^{\hat{k}-1}(1-p_i)^{n-\hat{k}} = \lim_{t\to\infty}\Pr(\omega = I, |i|_{n-1} = \hat{k}-1|g,t)$ by $p_i > 1/2$. Thus we obtain $\lim_{t\to\infty} EU_j(c|g) - EU_j(a|g) \leq -\underline{q}\binom{n-1}{\hat{k}-1}(1-p_i)^{\hat{k}-1}p_i^{n-\hat{k}} < 0$. Hence for any sufficiently large t, $EU_j(c|g) - EU_j(a|g) < 0$ for any juror $j \in N^t$. In other words, all jurors in N^t have strict incentives to vote for acquittal even if they observe private signal g, which establishes that informative voting is not an equilibrium.

In the case with $p_g < p_i$, an argument analogous to the above one establishes that, for any sufficiently large t, all jurors in N^t who observe private signal i have strict incentives to vote for conviction. Hence informative voting does not form an equilibrium in this case either. □

References

Austen-Smith, D., and J. S. Banks (1996): "Information Aggregation, Rationality, and the Condorcet Jury Theorem," *American Political Science Review*, 90, 34–45.

— , and T. J. Feddersen (2006): "Deliberation, Preference Uncertainty, and Voting Rules," *American Political Science Review*, 100, 209–217.

Coughlan, P. J. (2000): "In Defense of Unanimity Jury Verdicts: Mistrials, Communication, and Strategic Voting," *American Political Science Review*, 94, 375–393.

Duggan, J., and C. Martinelli (2001): "A Bayesian Model of Voting in Juries," *Games and Economic Behavior*, 37, 259–294.

Feddersen, T., and W. Pesendorfer (1998): "Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting," *American Political Science Review*, 92, 23–35.

Grofman, B., and S. Feld (1988): "Rousseau's General Will: A Condorcetian Perspective," *American Political Science Review*, 82, 567–576.

Guarnaschelli, S., R. D. McKelvey, and T. R. Palfrey (2000): "An Experimental Study of Jury Decision Rules," *American Political Science Review*, 94, 407–423.

Ladha, K. (1992): "The Condorcet Jury Theorem, Free Speech, and Correlated Votes," *American Journal of Political Science*, 36, 617–634.

McLennan, A. (1998): "Consequences of the Condorcet Jury Theorem for Beneficial Information Aggregation by Rational Agents," *American Political Science Review*, 92, 413–418.

Meirowitz, A. (2002): "Informative Voting and Codorcet Jury Theorems with a Continuum of Types," *Social Choice and Welfare*, 19, 219–236.

Wit, J. (1998): "Rational Choice and the Condorcet Jury Theorem," *Games and Economic Behavior*, 22, 364–376.

Young, P. H. (1988): "Condorcet's Theory of Voting," *American Political Science Review*, 82, 1231–1244.