# PUBLIC INSTITUTIONS AND PRIVATE INCENTIVES: THREE ESSAYS

Thesis by

Peter J. Coughlan

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1999
(Submitted June 7, 1999)

# ACKNOWLEDGMENTS

I would like to thank my colleagues at Caltech for many valuable insights and discussions. I am particularly indebted to Tom Palfrey, John Ledyard, Charles Plott, Richard McKelvey, and Simon Wilkie for providing invaluable guidance and support. Lastly, I am grateful to the Caltech Division of Humanities and Social Sciences, the National Science Foundation, and the Caltech Laboratory for Experimental Economics and Political Science for providing financial support for this research.

# TABLE OF CONTENTS

# Chapter I:

## Introduction

Public institutions are ubiquitous in economic environments of all types and sizes, and these institutions exert significant influence on the outcomes of innumerable economic interactions. Public institutions develop wherever a group of individuals, firms, or other economic entities possess some overriding common interest or social concern that is inadequately addressed in the absence of collective or government intervention. Public institutions therefore often evolve to promote the aggregate benefit of a society, group, or organization when individual members of such a collective can not be expected to efficiently coordinate actions for the common good.

In contrast to Adam Smith's notion of the "invisible hand" pushing society towards common goals while individuals pursue their own selfish interests, there are many economic environments in which private interests conflict with collective priorities. Such problems may arise from the presence of economic externalities, an inability of individuals to commit to future actions, an insufficiency of information, or excessive transaction costs in the absence of intervention. The development of collective institutions to address these conflicts between private and public interests naturally leads to two avenues of economic analysis. The first research approach asks the normative question, "What would be the optimal public institution in such an environment?", while the second approach asks the positive question, "What are the effects of a particular public institution on economic behavior and collective outcomes?" In this thesis, both types of questions are asked in the analysis of three separate public institutions.

In Chapter II, we investigate public institutions designed to efficiently allocate an excludable and congestible public good. In particular, we focus on an economic environment in which the public good is produced at constant returns to scale up to a maximum feasible level, and in which individuals have independent private valuations for the public good and congestion functions which adjust their consumption utility based on the consumption of others. The set of all interim efficient allocation rules in this asymmetric information environment is fully characterized using a Bayesian implementation approach. We find that the description of optimal allocation rules relies heavily upon the use of the concept of virtual valuation, which is a function of the true public good valuation, the probability distribution of valuations, and the welfare weighting function for each particular individual. In general, optimal exclusion in this environment requires that an individual be excluded from consumption if and only if his inclusion would lower the sum of included virtual valuations adjusted for congestion. In other words, if the negative congestion effect an individual's consumption creates is greater than the positive benefit that this individual gets from consuming the public good, then that individual must be excluded. The optimal public good production solution is then to produce the maximum feasible level of the public good whenever the sum of the included virtual valuations adjusted for congestion is greater than the cost of production. We further demonstrate that the conclusions of this analysis can be adapted to characterize the set of interim efficient mechanisms in several environments where special conditions exist, such as no exclusion, no congestion, complementarity of consumption, or identical congestion effects across individuals. In this last case of identical congestion functions, we find that the optimal set of consumers of the public good is all individuals whose virtual valuations are greater than or equal to some particular threshold, the value of which depends on both the individual congestion functions and the actual realization of individual valuations. This conclusion is analogous to existing results for allocation of an excludable but non-congestible public good.

In Chapter III, we explore the public institution of jury trials for determining the fate of a criminal defendant. It is a widely held belief among legal theorists that the requirement of unanimous jury verdicts in such trials reduces the likelihood of convicting an innocent defendant. This belief is, to a large extent, dependent upon the assumption that all jurors will vote non-strategically based on their own impression of the trial evidence. Recent literature, however, has drawn this assumption into question, and simple models of jury procedure have been constructed in which it is never a Nash equilibrium for all jurors to vote non-strategically under unanimity rule. Moreover, Nash equilibrium behavior in these models leads to higher probabilities of both convicting an innocent defendant and acquitting a guilty defendant under unanimity rule than under a wide variety of alternative voting rules, including simple majority rule. The present paper extends this research by adding minimal enhancements that we argue bring the existing models closer to actual jury procedures. In particular, we separately analyze the implications of (1) incorporating the possibility of mistrial and (2) allowing limited communication among jurors. Under each of these enhancements, we identify general conditions under which non-strategic voting is, in fact, a Nash equilibrium. We further demonstrate that under such equilibria, unanimous jury verdicts perform better than any alternative voting rule in terms of minimizing probability of trial error and maximizing expected utility, thus reversing the conclusions of the previous analysis.

Finally, in Chapter IV, we examine a different legal institution, namely the system for allocating legal costs among litigants in a civil lawsuit. The expanding volume of such lawsuits and the ballooning of legal expenditures in recent years has attracted the interest, concern, and even anger of the American public and politicians. These developments have led law makers to consider alternative legal fee allocation rules as methods for administering justice more efficiently. Under the traditional American rule, parties to a lawsuit must each pay their own legal expenses. One reform proposal is the English rule, under which the losing party must pay the prevailing party's attorney fees in addition to her own expenses. To evaluate the different effects of these two rules on litigant behavior and legal outcomes, we conduct a theoretical and experimental analysis of environments which can be interpreted as legal disputes in which the probability of winning a lawsuit is partially determined by the legal expenditures of the litigants and partially determined by the inherent merits of the case. We investigate decisions regarding trial expenditure and examine the effects of the two allocation rules on pretrial issues of suit and settlement. The data demonstrate that game theoretic equilibrium models produce good qualitative predictions of the relative institutional response to changes in the allocation rule and to differences in such parameters as case merit and lawyer productivity. In our most significant result, we find that the English rule produces significantly higher expenditure at trial than the American rule. On the other hand, the frequency of trial is significantly lower under the English rule. Combining these two effects, we find that average expenditure per legal dispute is higher under the English rule than under the American rule.

# Chapter II:

# Efficient Allocation of a Congestible and Excludable Public Good

## 1. Introduction

This paper addresses the classic economic problem of deciding whether or not to produce a public good and how to allocate the associated costs, which has been one of the primary focuses of research in the design of optimal allocation mechanisms. While the basic problem of production and cost allocation of a pure public good has been extensively explored, we know much less about the optimal solution in the presence of "impurities" such as exclusion and congestion. Therefore, this paper seeks to provide a characterization of optimal allocation mechanisms for the broader class of public goods that may be excludable and/or congestible.

In particular, we consider the problem of a group of individuals who must choose the level of a public good that is produced according to constant returns up to a maximum feasible level. The public good is assumed to be excludable, and therefore it must also be decided how much of the public good each individual will be permitted to consume. Lastly, the group must determine how to tax the individuals such that the total tax revenue covers the cost of producing the public good. Our analysis also allows for the possibility that consumption of the public good by an individual may create an externality that impacts the utility of other individuals. Such an externality will be generally referred to as a congestion effect; however, the externality created may be either positive or negative. Moreover, the effects of congestion may impact different members of the group differently, but the individual congestion functions are assumed to be common knowledge. Each individual has a particular "valuation" for the public good, which is equivalent to the individual's marginal rate of substitution, in the absence of congestion, between the public good and the private good tax payment. Each individual is assumed to know her own valuation for the public good, but not the valuations of the other group members. Adopting a Bayesian mechanism design approach, we assume that the prior probability distribution of each individual's valuation is common knowledge.

Examples of public goods that fit into the framework analyzed in this paper are numerous and varied. They include community facilities such as golf courses, swimming pools, parks, and libraries, as well as shared transportation resources including airports, bridges, and highways. Also fitting the description are telephone systems, computer networks, and any shared resources within a firm or organization. All of these shared goods are often excludable and they may exhibit consumption externalities, be they positive or negative.

The particular focus of this paper addresses two similar yet separate concerns. We seek to determine both which mechanisms are "optimal," in terms of maximizing some social welfare function, and which mechanisms are "efficient," in the sense that they are stable and unlikely to be abandoned for a more preferred mechanism. Fortunately, as we will discuss below, an existing result allows us to answer these two questions simultaneously.

Optimality of an allocation is often determined by measuring the value of some social welfare function in which individual utilities are each given some welfare weight, usually in a linear summation across individuals. The benchmark case gives equal welfare weight to all individuals, regardless of each individual's public good valuation or any other differentiating characteristic. An important consequence of such neutral welfare weighting schemes is that total social welfare is independent of cost allocation across individuals when utility is linear in the private good tax. Therefore, optimality in such cases is simply a measure of the production decision (and consumption decision, where applicable) but not the cost allocation decision. A planner may, however, be concerned with the allocation of cost across individuals. The planner may, for example, wish to treat individuals differently who value the public good differently or to use the private good tax to redistribute wealth in a particular way. Therefore, we adopt a more general concept of optimality in which the social welfare function may weight different types of individuals differently.

The question of efficiency, on the other hand, is concerned with whether a mechanism is sufficiently stable such that the group of individuals does not prefer an alternative feasible mechanism. Thus, to measure efficiency, we seek to use a concept appropriate for the asymmetric information framework that is analogous to Pareto efficiency in complete information environments. We therefore use an extension of Pareto efficiency referred to as interim efficiency, which is applicable at the interim stage of the mechanism, when each individual knows her own public good valuation (or type) but does not yet know the valuations of the other members of the group. An interim efficient mechanism is then an incentive compatible mechanism in which, in the absence of communication, it can not be common knowledge that there is another mechanism under which all individuals are better off (or at least as well off).

Fortunately, these two concepts of optimality and efficiency do not need to be investigated separately. Holmström and Myerson (1983) demonstrated that a mechanism is interim efficient if and only if there exist type-dependent welfare weights for which the mechanism maximizes the social welfare function across all feasible and incentive compatible mechanisms. Hence, the set of interim efficient mechanisms is equivalent to the set of all mechanisms which are optimal for some social welfare weights. For this reason, in our discussion we will often refer to a mechanism as optimal or efficient interchangeably.

The organization of the remainder of the paper is as follows. In section 2, the existing research is discussed as it relates to our particular problem, while our model of an excludable and congestible public good environment is presented in section 3. Section 4 is the key section of the paper, in which mechanisms and associated properties are discussed, the planner's optimization problem is described, the constrained maximization is analyzed, and the characterization theorem is presented. Section 5 presents discussion and interpretation of the theorem as well as an example problem and solution, while section 6 applies the general results to several special cases. Lastly, section 7 provides conclusions and extensions.

## 2. Literature Review

As mentioned previously, there is an extensive literature on public good allocation problems. One of the most significant intellectual events in this line of research was the development of the class of so-called "demand revealing" or "pivotal" mechanisms by Clarke (1971) and Groves (1973). These mechanisms are the public good analogue of the second price auction of Vickrey (1961), and they implement in dominant strategies allocation rules which maximize the sum of individual utilities (i.e., a social welfare function with neutral welfare weights). One significant problem with these mechanisms, however, is that the associated tax functions do not balance the budget in terms of exactly covering production costs, but rather may produce either a surplus or deficit. In addition, these mechanisms are not immune from manipulation by coalitions, as demonstrated by Bennett and Conn (1977).

One classic dominant strategy mechanism for implementing public goods that does balance the budget and is coalition strategyproof is the conservative equal costs mechanism. Under this mechanism, the level of the public good chosen is the smallest demand of all individuals, and the costs of producing this amount is shared equally among all members of the group. While this mechanism uniquely satisfies several significant normative criteria (Moulin 1994), there is significant room for improvement in efficiency. A Pareto superior mechanism with similar normative characteristics which is applicable to the allocation of excludable public goods is the serial cost sharing mechanism developed by Moulin and Shenker (1992). This mechanism is characterized by two properties: (1) cost shares depend anonymously upon demands, and (2) an agent's cost share is independent of demands higher than her own. For the case of an indivisible and excludable public good, Deb and Razzolini (1994) describe "auction like mechanisms" that are roughly equivalent to serial cost sharing in this context. None of these dominant strategy mechanisms were designed to accommodate congestible public goods, however, and their basic normative properties fail to hold true in this presence of congestion. On the other hand, it is important to note that when the congestion effects satisfy certain regularity assumptions, an analogue for mechanisms such as serial cost sharing can be developed which are strategyproof and possess many of the same normative characteristics.

Much of the research on implementation of allocation rules in Nash equilibrium has been inspired by the work of Groves and Ledyard (1977). These authors developed an incentive scheme in which the Nash equilibrium outcomes are all Pareto efficient for economies with any number of private and public goods and preferences that are only minimally restricted. In addition, Hurwicz (1979) developed an allocation rule whose Nash equilibrium allocations coincide with the Lindahl equilibria of the economy, while Walker (1981), Tian (1989), and Peleg (1996) later obtained similar results employing simpler mechanisms. The optimality of the Lindahl solution is not always apparent, however, given that it may violate individual rationality and coalition strategyproofness in the absence of constant returns to scale. Corchon and Wilkie (1996), therefore, identify a simple market game which implements the ratio equilibrium, perhaps a more appealing solution, in both Nash and strong equilibria. Also Nash implementable are the dynamic "MDP" public good allocation procedures developed by Malinvaud (1972), Drèze and de la Vallée Poussin (1971), and Tideman (1972). In these procedures, consumers report their preferences at each instant in time and the planner uses this information to determine the level of the public good and the individual tax payments. The appealing normative characteristic of these MDP procedures is that the equilibrium allocations converge to a Pareto optimum over time.

In the asymmetric information environment in which the concept of Bayesian equilibrium is applied, one of the most significant results is that of d'Aspremont and Gérard-Varet (1979). They demonstrated that, in the case of a pure public good and quasi-linear utility, there exist Bayesian implementable allocation rules that maximize the sum of individual utilities while also exactly covering costs of production. Laffont and Maskin (1979) fully characterized the class of such Bayesian implementable rules and highlighted its close connection to the class of demand revealing mechanisms.

When individual valuations for a pure public good produced at constant unit cost are identically distributed and can only take two possible values, the set of interim efficient mechanisms was fully characterized by Ledyard and Palfrey (1994). In this case, efficient allocation rules have the property that the public good is produced if and only if the number of high valuation types exceeds some threshold which depends both on the welfare weights and the distribution of types. Ledyard and Palfrey (1999) extended this line of research by also characterizing the interim efficient allocation rules when the valuations of the pure public good can take on any value within a given interval for each individual. In this case, they find that it is optimal to produce the public good if and only if the sum of "virtual valuations" across individuals exceeds the production cost, where the virtual valuation for each individual is a function of the true valuation, the distribution of valuations, and the welfare weighting function for that individual. It is this work by Ledyard and Palfrey that most directly inspired the present paper and their analysis approach is followed quite closely in developing the characterization theorem in section 4. Cornelli (1996) conducts an analysis in which the public good is excludable and an individual rationality constraint is also imposed. Although the focus of the Cornelli work is primarily on the profit maximizing mechanism for a monopolist, the research is also applied to the maximization of a neutral social welfare function in which case the public good, when produced, is optimally provided to any individual whose virtual valuation is positive, where the virtual valuation is a function of both the true valuation and the distribution of valuations for a given individual. The optimal production decision in this case is to produce the public good if and only if the sum of the positive virtual valuations exceeds the production cost. The present paper directly extends the work of Cornelli (as well as Ledyard and Palfrey) by allowing the public good to be congestible as well as excludable, allowing the welfare weights to be type-dependent, and not imposing an individual rationality constraint.

Before proceeding, it is important to discuss the current literature on the general issue of congestion. While the existing research which takes a mechanism design approach to the problem of congestion in public goods is quite limited, there is extensive investigation of specific congestion issues in several areas of economics. In the area of club goods and local public goods, congestion is sometimes incorporated into cooperative game theoretic models which are used to identify optimal club size and the core distribution of individuals among multiple clubs. Prominent examples of such research includes work by Pauly (1967 and 1970) and Sorenson, Tschirhart, and Whinston (1978). Other researchers in this area, such as Scotchmer and Wooders (1987), have included congestion effects in general equilibrium models of club economies. A separate field of investigation in which congestion is a factor is the field of environmental economics, where the issue of dissipation of common pool resources is often addressed. This line of research usually models the problem as one of a congestible but not excludable public good, and includes the work of Clark (1980), Gardner, Ostrom, and Walker (1990 and 1992), and Ito, Saijo, and Une (1995). There is also frequent concern with the issue of congestion in the field of transportation economics, in which common areas of investigation include optimal tolls and efficient road usage. Research on congestion in this field usually deals with specialized issues such as stochastic congestion, multiple competing public goods (roads), and the congestion relief effects of transportation investment. Else (1981), Arnott, de Palma, and Lindsey (1992), and Verhoef et al. (1996) are representative examples of this transportation literature, while Berglas and Pines (1981) have attempted to synthesize the transportation models with the previously mentioned literature on club goods and local public goods.

The present paper significantly contributes to research on issues of congestion in all of the described fields. The model we present in the next section is sufficiently generalizable to apply to the scenarios commonly investigated in the club goods, environmental economics, and transportation economics literature. Moreover, this paper enhances our understanding in each of these areas by incorporating important factors often not present in prevailing models. For example, much of the club goods and local publics literature downplays issues of asymmetric information and incentive compatibility, focusing instead on characterizing, rather than implementing, optimal solutions. Research in all of these fields also frequently centers on scenarios with homogeneous individuals, rather than individuals who may have different preferences or be affected differently by congestion. Other common limitations of the existing congestion literature that are not present in this paper include the absence of possible exclusion, the constraint of a single tax or financial incentive for all individuals, and the treatment of certain choice variables as probabilistic.

## 3. The Model

Consider an economy with a single private good and a single excludable and congestible public good. The population in this economy is given by a set of individuals $N = \{1,2,\dots,n\}$. These individuals must collectively determine three variables: (1) the total quantity of the public good to produce, (2) the proportion of the public good that each individual will be allowed to consume, and (3) how the cost of producing the public good will be shared among the individuals.

### 3.1 Public Good Production, Consumption, and Cost Allocation

Let the quantity of public good produced be given by x, where $x \in [0,1]$. The public good is produced using a constant returns to scale technology with unit cost K. Thus the cost of producing a quantity x of the public good is equal to Kx. Because of this linear cost function, the optimal level of public good production will always be either 0 or 1. Therefore, the production decision is equivalent to deciding whether or not to produce a discrete public good.

Because the public good is excludable, it must also be determined how much of the public good each individual will be allowed to consume. Let $p_i$ be the proportion of the public good consumed by individual i, where $p_i \in [0,1]$. We will denote by $p=(p_1,p_2,\ldots,p_n)$ the vector of public good consumption quantities for all individuals.

To cover the cost of producing the public good, each individual may be charged a "tax." For each $i \in N$, let $t_i \in \Re$ denote the tax payment from individual i. These taxes are in units of consumption of the private good, and we will denote by $t=(t_1,t_2,\ldots,t_n)$ the vector of taxes for all individuals. Therefore, in order for the public good production costs to be covered by the tax payments, we must have:

$$\sum_{i=1}^{n} t_i \text{ ? } Kx$$

### 3.2 Preferences

Individual preferences are assumed to be quasilinear in consumption of the public good and the tax payment. The utility to individual i for allocation (x,p,t) is given by $v_i \cdot x \cdot p_i \cdot c_i(p) - t_i$.

The value $v_i$ can be interpreted as individual i's type, and represents her valuation of the public good. We refer to $v_i$ as player i's "value" and assume that each individual knows her own value but does not know the values of the other individuals. We will denote by v = $(v_1,v_2,\ldots,v_n)$ the vector of values for all individuals. These values are assumed to be independently distributed, with the cumulative distribution function for $v_i$ being $F_i(\cdot)$ with support $V_i = [\underline{v}_i, \overline{v}_i]$. Note that $\underline{v}_i < 0$ is allowed and therefore negative values are possible. Let $F(\cdot) = F_1(\cdot)F_2(\cdot)\cdots F_n(\cdot)$ be the joint distribution function for v with support $V_1 \times V_2 \times \cdots \times V_n$. We assume that all distribution functions are common knowledge and that each $F_i(\cdot)$ has an associated density function, $f_i(\cdot)$, which is continuous and positive on $V_i$.

The function $c_i(p)$ represents the congestion function for individual i. This function adjusts individual i's utility of consuming the public good based on any externality imposed by the consumption of others, which is given by the vector p. For each $i \in N$, we assume that the congestion function $c_i(p)$ is common knowledge.

Note that the total effect of congestion on individual utility may be negative, neutral, or positive. To illustrate this, suppose individual i has a non-negative valuation for the public good $v_i \geq 0$. Then if $c_i(p) < 1$, we have a situation of congestion or crowding in which individual i's utility is reduced as a result of the consumption vector q. If $c_i(p) = 1$, on the other hand, we have a case of no congestion in which individual i's utility is unaffected by the consumption of others. Lastly, it may be the case that $c_i(p) > 1$, which represents a situation of camaraderie or complementarity, in which individual i enjoys a positive externality from sharing consumption of the public good with other individuals.

Also recognize that, by the nature of the utility formulation, the consumption of others is assumed to impact an individual's consumption utility in a multiplicative fashion. An alternative approach is to assume that congestion has an additive impact on utility, with individual utility being formulated something like $v_i \cdot x \cdot p_i - c_i(p) - t_i$, for example. While this additive formulation eliminates some computational complexity, it is not as appealing intellectually in that it does not seem to capture the full effect of the consumption externalities we are modeling. In particular, it seems that congestion effects should affect not only overall utility, but also the marginal utility of each unit of public good consumption. For example, crowding at your community pool affects your utility for each visit to the pool, not just your overall utility of pool membership. This important marginal utility effect is, of course, not present in the additive formulation. Nonetheless, for completeness of understanding, the additive congestion formulation should be investigated in future research.

## 4. Optimal Allocation Mechanisms

In this section, we identify the requirements for an allocation mechanism to be optimal in this framework, setup the optimization problem, and conclude with a theorem providing a full characterization of the set of optimal mechanisms. It should be noted that the analytical approach in this section borrows heavily from Ledyard and Palfrey (1999).

A *mechanism* consists of a message space for each individual and an outcome function mapping message profiles into the set of feasible allocations. The revelation principle tells us that the allocation properties of any optimal mechanism can be duplicated by a direct mechanism in which each individual simply reports a type or value. Therefore, we can restrict our attention to mechanisms in which the message space for individual i is simply the set of possible values for individual i, $V_i$, and the joint message space for all individuals is the set of possible vectors of values, V. A direct mechanism is a function $\eta(v) = (x(v),p(v),t(v))$, where x(v) is the total quantity of the public good produced at profile v, $p_i(v)$ is the proportion of the public good consumed by individual i at profile v, and $t_i(v)$ is the private good tax of individual i at profile v.

### 4.1 Interim Utility

The *interim utility* for individual i of report (or message) $w_i$ given value $v_i$ is denoted $u_i(v_i, w_i)$ and it is the expected utility for individual i when she has a value of $v_i$ and reports a value of $w_i$ while all other individuals truthfully report their values. Thus, $u_i(v_i, w_i)$ is given by:

$$u_i(v_i, w_i) = \int_{V_{-i}} \left[ v_i x(w_i, v_{-i}) p_i(w_i, v_{-i}) - c_i(p(w_i, v_{-i})) - t_i(w_i, v_{-i}) \right] dF_{-i}(v_{-i})$$

where

$$F_{-i}(v_{-i}) = F_1(v_1)F_2(v_2)\cdots F_{i-1}(v_{i-1})F_{i+1}(v_{i+1})\cdots F_n(v_n)$$

$$V_{-i} = V_1 \times V_2 \times \cdots \times V_{i-1} \times V_{i+1} \times \cdots \times V_n$$

Let $\bar{u}_i(v_i) = u_i(v_i, v_i)$. In other words, $\bar{u}_i(v_i)$ is the expected utility for individual i when she truthfully reports her value while all other individuals also truthfully report their values. We will refer to $\bar{u}_i(v_i)$ as the *truthful interim utility* for individual i. Formally, $\bar{u}_i(v_i)$ is given by:

$$\bar{u}_i(v_i) = \int_{V_{-i}} [v_i x(v) p_i(v) c_i(p(v)) - t_i(v)] dF_{-i}(v_{-i})$$

To allow us to better interpret these interim utilities, we define:

$$P_i(w_i) = \int_{V_{-i}} x(w_i, v_{-i}) p_i(w_i, v_{-i}) c_i(p(w_i, v_{-i})) dF_{-i}(v_{-i})$$
$$T_i(w_i) = \int_{V_{-i}} t_i(w_i, v_{-i}) dF_{-i}(v_{-i})$$

Thus, $P_i(w_i)$ is the expected public good consumption for individual i, adjusted for congestion, when she reports a value of $w_i$ and all others report truthfully. Similarly, $T_i(w_i)$ is the expected tax payment for individual i when she reports a value of $w_i$ and all others report truthfully. Using these simplifications allows us to write $u_i(v_i, w_i) = v_i P_i(w_i) - T_i(w_i)$ and $\bar{u}_i(v_i) = v_i P_i(v_i) - T_i(v_i)$.

### 4.2 Feasibility, Incentive Compatibility, & Interim Efficiency

There are three primary restrictions on optimal allocation mechanisms that we will impose in our analysis. The first of these restrictions is *feasibility*, which places bounds on the range of the mechanism's outcome function. In particular, a feasible direct mechanism $\eta$ is a function satisfying:

$$\eta: V \rightarrow \Omega \text{ where } \Omega = \left\{ (x,p,t) \in [0,1]^{n+1} \times \Re^n \,\middle|\, \sum_{i=1}^{n} t_i \geq Kx \right\}$$

The second fundamental restriction on an optimal allocation mechanism is *incentive compatibility*. This restriction requires that it be a Bayesian equilibrium of the mechanism for all individuals to truthfully report their value. This means that $\eta=(x,p,t)$ is incentive compatible if and only if:

$$\bar{u}_i(v_i) \geq u_i(v_i, w_i) \; \forall i \in N, \; \forall v_i, w_i \in V_i$$

The set of incentive compatible mechanisms for this class of problems can readily be characterized in terms of derivatives of the interim utility functions. The general characterization conditions are given by the following lemma.

Lemma (Rochet 1987): If $u_i(v_i, w_i)$ is linear with respect to $v_i$ and continuously differentiable with respect to $w_i$, then $\eta$ is incentive compatible if and only if:

$$\text{(i)} \quad \frac{\partial \bar{u}_i(v_i)}{\partial v_i} = \left. \frac{\partial u_i(v_i, w_i)}{\partial v_i} \right|_{w_i = v_i}$$

(ii) $\bar{u}_i(v_i)$ is convex in $v_i$.

The lemma thus identifies two conditions which are necessary and sufficient for incentive compatibility. The first is an envelope condition which requires that the total derivative with respect to value of the truthful interim utility be equal to the partial derivative with respect to value of the general interim utility evaluated at a truthful report. The second condition is a second-order restriction which requires that the truthful interim utility function be convex with respect to an individual's value.

Applying this lemma to our particular problem, we have:

$$\frac{f u_i(v_i, w_i)}{f v_i}\bigg|_{w_i=v_i} = \frac{f}{f v_i}\left(v_i P_i(w_i) - T_i(w_i)\right)\bigg|_{w_i=v_i} = P_i(w_i)\big|_{w_i=v_i} = P_i(v_i)$$

Thus, condition (i) in the lemma, when applied to our model, becomes:

$$\frac{f \bar{u}_i(v_i)}{f v_i} = P_i(v_i)$$

Condition (ii) of the lemma requires that $\bar{u}_i(v_i)$ be convex in $v_i$ or, alternatively, that:

$$\frac{f^2 \bar{u}_i(v_i)}{f v_i^2} \; ? \; 0, \; \forall v_i \, L \, V_i$$

$$? \; \frac{f}{f v_i}\left[\frac{f \bar{u}_i(v_i)}{f v_i}\right] ? \; 0, \; \forall v_i \, L \, V_i$$

$$? \; \frac{f}{f v_i} P_i(v_i) \; ? \; 0, \; \forall v_i \, L \, V_i$$

$$? \; P_i'(v_i) \; ? \; 0, \; \forall v_i \, L \, V_i$$

Thus, in our model, a direct mechanism $\eta$ is incentive compatible if and only if:

(i) $\frac{f \bar{u}_i(v_i)}{f v_i} = P_i(v_i)$

(ii) $P_i'(v_i) \; ? \; 0 \quad \forall i \in N, \forall v_i \in V_i$.

The final restriction we will place on an optimal allocation mechanism is *interim efficiency*. An allocation rule is interim efficient if (a) it is both feasible and incentive compatible, and (b) there exists no other feasible and incentive compatible allocation rule that makes a positive measure of types better off without also making a positive measure of types worse off. We represent this restriction by requiring that an optimal allocation mechanism be the solution to the following maximization problem.

For each $i \in N$, let $\lambda_i : V_i \rightarrow \Re_{++}$ be a welfare weighting function mapping values or types for individual i into the positive real line, such that $\lambda_i(v_i) > 0$ is the welfare weight assigned to type $v_i$ of individual i. Define $\lambda = (\lambda_1, \lambda_2, \ldots \lambda_n)$. Then an allocation mechanism $\eta$ is interim efficient if and only if there exists a $\lambda$ such that $\eta$ maximizes

$$\sum_{i=1}^{n} \int_{V_i} \lambda_i(v_i)\bar{u}_i(v_i) f_i(v_i) dv_i$$

over the set of all feasible and incentive compatible mechanisms.

Note that we must place certain restrictions on the welfare weighting functions to guarantee that a solution to this maximization problem is well defined. To facilitate discussion of these restrictions, let $\bar{\lambda}_i$ denote the expected welfare weight for individual i and be given by:

$$\bar{\lambda}_i = \int_{V_i} \lambda_i(v_i) f_i(v_i) dv_i \ , \ \forall i \in N$$

As discussed by Ledyard and Palfrey (1999), a solution to the maximization problem only exists when we have $\bar{\lambda}_i = \bar{\lambda}_j = \bar{\lambda} < \infty$ for all $i,j \in N$. Moreover, without loss of generality, we can normalize the welfare weights to satisfy $\bar{\lambda} = 1$. We will use this normalization in the analysis and discussion that follows.

### 4.3 The Optimization Problem

Having now identified the conditions that must be satisfied by all optimal allocation mechanisms, we can represent such interim efficient mechanisms as the solution to a constrained optimization problem. In particular a direct mechanism $\eta = (x,p,t)$ is interim efficient if and only if there exists a $\lambda >> 0$ such that $(x,p,t)$ solves:

$$\max \sum_{i=1}^{n} \int_{V_i} \lambda_i(v_i)\bar{u}_i(v_i)f_i(v_i)dv_i$$

subject to:

$$\text{(i)} \quad \frac{\partial \bar{u}_i(v_i)}{\partial v_i} = P_i(v_i)$$

$$\text{(ii)} \quad \sum_{i=1}^{n} t_i(v) = Kx(v)$$

$$\text{(iii)} \quad 0 \leq x(v) \leq 1 \quad \forall v \in V$$

$$\text{(iv)} \quad 0 \leq p_i(v) \leq 1 \quad \forall i \in N,\ \forall v \in V$$

$$\text{(v)} \quad P_i'(v_i)\ ?\ 0 \quad \forall i \in N,\ \forall v_i \in V_i$$

Employing the approach of Mirrlees (1971) and Wilson (1993), we construct the Lagrangian equivalent problem:

$$\max_{\eta=(x,p,t)} \min_{\psi,\delta} \Bigg[ \sum_{i=1}^{n} \int_{V_i} \lambda_i(v_i)\bar{u}_i(v_i)f_i(v_i)dv_i + \sum_{i=1}^{n} \int_{V_i} \psi_i(v_i)\left(\frac{\partial \bar{u}_i(v_i)}{\partial v_i} - P_i(v_i)\right)dv_i + \int_V \delta(v)\left(\sum_{i=1}^{n} t_i(v) - Kx(v)\right)dv \Bigg] \qquad (1)$$

where $\psi_1, \psi_2, \ldots, \psi_n$ are multipliers for the first order incentive compatibility constraints, and $\delta$ is the multiplier for first order feasibility constraint. Note that we still have the requirements that $0 \leq x(v) \leq 1$ for all $v \in V$, that $0 \leq p_i(v) \leq 1$ for all $i \in N$ and all $v \in V$, and that $P_i'(v_i)$ ? 0 for all $i \in N$ and all $v_i \in V_i$.

Our first step in solving this optimization problem is to employ several simplifying conversions. For example, applying integration by parts we find that:

$$
{}_{V_i}\psi_i(v_i)\frac{f\bar{u}_i(v_i)}{f v_i}dv_i = - {}_{V_i}\psi_i'(v_i)\bar{u}_i(v_i)dv_i + \left[\psi_i(v_i)\bar{u}_i(v_i)\right]_{\underline{v}_i}^{\bar{v}_i}
$$

$$
= - {}_{V_i}\psi_i'(v_i)\bar{u}_i(v_i)dv_i + \psi_i(\bar{v}_i)\bar{u}_i(\bar{v}_i) - \psi_i(\underline{v}_i)\bar{u}_i(\underline{v}_i)
$$

This equivalence allows us to rewrite (1) as:

$$
\max_{\eta=(x,p,t)} \min_{\psi,\delta} \left[ \sum_{i=1}^{n} \int_{V_i} \left\{ \left(\lambda_i(v_i)f_i(v_i) - \psi_i'(v_i)\right)\bar{u}_i(v_i) - \psi_i(v_i)P_i(v_i) \right\} dv_i \right.
$$

$$
\left. + \int_V \delta(v)\left(\sum_{i=1}^{n} t_i(v) - Kx(v)\right)dv + \sum_{i=1}^{n}\left(\psi_i(\bar{v}_i)\bar{u}_i(\bar{v}_i) - \psi_i(\underline{v}_i)\bar{u}_i(\underline{v}_i)\right) \right] \tag{2}
$$

Note that we will demonstrate below that the last summation in (1) vanishes in the optimal solution. To maintain accuracy, however, we preserve it in our description of the problem. Now to further reduce the objective function, recall that

$$
\bar{u}_i(v_i) = {}_{V_{-i}}\left[v_i x(v) p_i(v) c_i(p(v)) - t_i(v)\right]dF_{-i}(v_{-i})
$$

This gives us:

$$
\begin{aligned}
&\int_{V_i} \left(\lambda_i(v_i) f_i(v_i) - \psi_i'(v_i)\right) \pi_i(v_i)\, dv_i \\
&= \int_{V_i} \left(\lambda_i(v_i) f_i(v_i) - \psi_i'(v_i)\right) \left\{ \int_{V_{-i}} \left[ v_i x(v) p_i(v) c_i(p(v)) - t_i(v) \right] dF_{-i}(v_{-i}) \right\} dv_i \\
&= \int_{V_{-i}} \int_{V_i} \left(\lambda_i(v_i) f_i(v_i) - \psi_i'(v_i)\right) \left(v_i x(v) p_i(v) c_i(p(v)) - t_i(v)\right) dv_i\, dF_{-i}(v_{-i}) \\
&= \int_V \left( \lambda_i(v_i) - \frac{\psi_i'(v_i)}{f_i(v_i)} \right) \left(v_i x(v) p_i(v) c_i(p(v)) - t_i(v)\right) dF(v)
\end{aligned}
\tag{3}
$$

Also recall that $P_i(v_i) = \int_{V_{-i}} x(v) p_i(v) c_i(p(v)) dF_{-i}(v_{-i})$. This gives us:

$$
\begin{aligned}
\int_{V_i} \psi_i(v_i) P_i(v_i)\, dv_i &= \int_{V_i} \psi_i(v_i) \int_{V_{-i}} x(v) p_i(v) c_i(p(v)) dF_{-i}(v_{-i})\, dv_i \\
&= \int_V \frac{\psi_i(v_i)}{f_i(v_i)} x(v) p_i(v) c_i(p(v)) dF(v)
\end{aligned}
\tag{4}
$$

We can also write

$$
\int_V \delta(v) \left[ \sum_{i=1}^n t_i(v) - Kx(v) \right] dv = \int_V \gamma(v) \left[ \sum_{i=1}^n t_i(v) - Kx(v) \right] dF(v) \tag{5}
$$

where

$$
\gamma(v) = \frac{\delta(v)}{f(v)} = \frac{\delta(v_1, v_2, \ldots, v_n)}{f_1(v_1) f_2(v_2) \cdots f_n(v_n)}
$$

Substituting equations (3), (4), and (5) into (2) converts the maximization problem to:

$$
\begin{aligned}
\max_{\eta=(x,p,t)} \min_{\psi,\gamma} \int_V \Bigg\{ \sum_{i=1}^n \Bigg[ &\left( \lambda_i(v_i) - \frac{\psi_i'(v_i)}{f_i(v_i)} \right) \left(v_i x(v) p_i(v) c_i(p(v)) - t_i(v)\right) \\
&- \frac{\psi_i(v_i)}{f_i(v_i)} x(v) p_i(v) c_i(p(v)) \Bigg] + \gamma(v) \left[ \sum_{i=1}^n t_i(v) - Kx(v) \right] \Bigg\} dF(v)
\end{aligned}
$$

$$+\sum_{i=1}^{n}\left(\psi_i(\overline{v}_i)\overline{u}_i(\overline{v}_i)-\psi_i(\underline{v}_i)\overline{u}_i(\underline{v}_i)\right)$$

or alternatively:

$$\max_{\eta=(x,p,t)}\ \min_{\psi,\gamma}\int_V\Bigg[x(v)\sum_{i=1}^{n}\left(\frac{v_i\lambda_i(v_i)f_i(v_i)-v_i\psi_i'(v_i)-\psi_i(v_i)}{f_i(v_i)}\right)p_i(v)c_i(p(v))$$
$$-\sum_{i=1}^{n}\left(\lambda_i(v_i)-\frac{\psi_i'(v_i)}{f_i(v_i)}\right)t_i(v)+\gamma(v)\left(\sum_{i=1}^{n}t_i(v)-Kx(v)\right)\Bigg]dF(v) \tag{6}$$
$$+\sum_{i=1}^{n}\left(\psi_i(\overline{v}_i)\overline{u}_i(\overline{v}_i)-\psi_i(\underline{v}_i)\overline{u}_i(\underline{v}_i)\right)$$

## 4.4 First Order Conditions

To address this optimization problem, we start by calculating the first order necessary conditions on the variables t, x, p, $\gamma$, and $\psi$. In doing so, it is important to recognize that expansion of the final summation in the objective function gives us:

$$\sum_{i=1}^{n}\left(\psi_i(\overline{v}_i)\overline{u}_i(\overline{v}_i)-\psi_i(\underline{v}_i)\overline{u}_i(\underline{v}_i)\right) = \sum_{i=1}^{n}\Big[\psi_i(v_i)\big(v_iP_i(v_i)-T_i(v_i)\big)\Big]_{\underline{v}_i}^{\overline{v}_i}$$
$$= \sum_{i=1}^{n}\left[\psi_i(v_i)\int_{V_{-i}}\big(v_ix(v)p_i(v)c_i(p(v))-t_i(v)\big)dF_{-i}(v_{-i})\right]_{\underline{v}_i}^{\overline{v}_i}$$

Thus, the boundary values of the functions p, t, and $\psi$ all show up in this final summation and therefore must be considered when calculating first order conditions for these choice variables. Fortunately, our analysis is greatly simplified because this final summation vanishes in the optimal solution.

To demonstrate this, we calculate the first order condition on the tax function $t_i(v)$. Differentiation of the objective function (6) with respect to $t_i(v)$ give us the following necessary condition for $\underline{v}_i < v_i < \bar{v}_i$ :

$$\frac{\psi_i'(v_i)}{f_i(v_i)} - \lambda_i(v_i) + \gamma(v) = 0 \quad \forall i \in N \tag{7}$$

At the boundary points $\underline{v}_i$ and $\bar{v}_i$, on the other hand, differentiation with respect to $t_i(v)$ give us:

$$\frac{\psi_i'(\underline{v}_i)}{f_i(\underline{v}_i)} - \lambda_i(\underline{v}_i) + \gamma(\underline{v}_i, v_{-i}) - \psi_i(\underline{v}_i) = 0 \quad \forall i \in N \tag{8}$$

$$\frac{\psi_i'(\bar{v}_i)}{f_i(\bar{v}_i)} - \lambda_i(\bar{v}_i) + \gamma(\bar{v}_i, v_{-i}) - \psi_i(\bar{v}_i) = 0 \quad \forall i \in N \tag{9}$$

Because the functions $\psi_i'(v_i)$, $f_i(v_i)$, $\lambda_i(v_i)$, and $\gamma(v)$ are all continuous in $v_i$, it must be the case that (7) holds, not only for values in the interior of $V_i$, but also at the boundary points $\underline{v}_i$ and $\bar{v}_i$. Therefore, equations (8) and (9) imply that $\psi_i(\underline{v}_i) = \psi_i(\bar{v}_i) = 0$ for all $i \in N$. Thus, the summation

$$\sum_{i=1}^{n} \left( \psi_i(\bar{v}_i)\bar{u}_i(\bar{v}_i) - \psi_i(\underline{v}_i)\bar{u}_i(\underline{v}_i) \right)$$

will vanish in the optimal solution, and we therefore suppress it in the calculation of the remaining first-order conditions.

To reduce notation in the remaining calculations, we define the function $\omega_i$: $V_i \rightarrow \Re$ as follows:

$$\omega_i(v_i) = \frac{v_i\lambda_i(v_i)f_i(v_i) - v_i\psi_i'(v_i) - \psi_i(v_i)}{f_i(v_i)}.$$

This function will be very significant in characterizing the optimal allocation rule, and it will be discussed in more detail later in the paper.

The next choice variable we address is the level of public good production, x(v), for which we have the feasibility constraint $x(v) \in [0,1]$ for all $v \in V$. With this constraint and the definition of $\omega_i(v_i)$ in mind, differentiation of the objective function in (6) with respect to x(v) yields:

$$\begin{aligned}
&\sum_{i=1}^{n}\omega_i(v_i)p_i(v)c_i(p(v)) - \gamma(v)K \geq 0 \quad \text{if } x(v) = 1 \\
&\sum_{i=1}^{n}\omega_i(v_i)p_i(v)c_i(p(v)) - \gamma(v)K = 0 \quad \text{if } x(v) \in (0,1) \\
&\sum_{i=1}^{n}\omega_i(v_i)p_i(v)c_i(p(v)) - \gamma(v)K \leq 0 \quad \text{if } x(v) = 0
\end{aligned} \tag{10}$$

For the consumption function, p(v), we also have a feasibility constraint that requires $p_i(v) \in [0,1]$ for all $i \in N$ and all $v \in V$. Thus, differentiation of the objective function with respect to $p_i(v)$ yields:

$$\begin{aligned}
&x(v)\left[\omega_i(v)c_i(p(v)) + \sum_{j=1}^{n}\omega_j(v)p_j(v)\frac{\partial c_j(p(v))}{\partial p_i(v)}\right] \geq 0 \quad \text{if } p_i(v) = 1 \\
&x(v)\left[\omega_i(v)c_i(p(v)) + \sum_{j=1}^{n}\omega_j(v)p_j(v)\frac{\partial c_j(p(v))}{\partial p_i(v)}\right] = 0 \quad \text{if } p_i(v) \in (0,1) \\
&x(v)\left[\omega_i(v)c_i(p(v)) + \sum_{j=1}^{n}\omega_j(v)p_j(v)\frac{\partial c_j(p(v))}{\partial p_i(v)}\right] \leq 0 \quad \text{if } p_i(v) = 0
\end{aligned} \tag{11}$$

Differentiation of the objective function with respect to the budget constraint multiplier, $\gamma_i(v)$, produces the following first order condition:

$$\sum_{i=1}^{n} t_i(v) - Kx(v) = 0$$

$$\Rightarrow \sum_{i=1}^{n} t_i(v) = Kx(v) \tag{12}$$

Finally, let us consider the first order conditions on the functions $\psi_i(v_i)$ for i = 1,2,...,n, which are the multipliers on the incentive compatibility constraints. Note that $\psi_i(v_i)$is a choice variable in the optimization of an integral objective function involving both $\psi_i(v_i)$ and $\psi_i'(v_i)$. Therefore, we have a calculus of variations problem in which the first-order necessary condition for an optimum is given by the following Euler equation:

$$\frac{\partial L}{\partial \psi_i(v_i)} = \frac{\partial}{\partial v_i}\frac{\partial L}{\partial \psi_i'(v_i)}$$

where *L* denote the Lagrangian objective function (excluding the final summation).

The partial derivative of the objective function in (6) with respect to $\psi_i(v_i)$ is given by:

$$\begin{aligned}
\frac{\partial L}{\partial \psi_i(v_i)} &= \frac{\partial}{\partial \psi_i(v_i)}\left[\int_V \frac{-\psi_i(v_i)}{f_i(v_i)} x(v)p_i(v)c_i(p(v))dF(v)\right] \\
&= \frac{\partial}{\partial \psi_i(v_i)}\left[-\psi_i(v_i)\int_{V_{-i}} x(v)p_i(v)c_i(p(v))dF_{-i}(v_{-i})\right] \\
&= -\int_{V_{-i}} x(v)p_i(v)c_i(p(v))dF_{-i}(v_{-i}) \\
&= -P_i(v_i)
\end{aligned}$$

On the other hand, the partial derivative of the objective function with respect to $\psi_i'(v_i)$ is given by:

$$\begin{aligned}
\frac{fL}{f\psi_i'(v_i)} &= \frac{\partial}{\partial\psi_i'(v_i)}\left[\int_V\left(\frac{\psi_i'(v_i)}{f_i(v_i)}t_i(v)-\frac{v_i\psi_i'(v_i)}{f_i(v_i)}x(v)p_i(v)c_i(p(v))\right)dF(v)\right] \\
&= \frac{\partial}{\partial\psi_i'(v_i)}\left[\psi_i'(v_i)\left(\int_{V_{-i}}t_i(v)dF_{-i}(v_{-i})-v_i\int_{V_{-i}}x(v)p_i(v)c_i(p(v))dF_{-i}(v_{-i})\right)\right] \\
&= \int_{V_{-i}}t_i(v)dF_{-i}(v_{-i})-v_i\int_{V_{-i}}x(v)p_i(v)c_i(p(v))dF_{-i}(v_{-i}) \\
&= T_i(v_i)-v_iP_i(v_i) = -\bar{u}_i(v_i)
\end{aligned}$$

Thus, our final first order necessary condition is:

$$\begin{aligned}
\frac{fL}{f\psi_i(v_i)} &= \frac{f}{fv_i}\frac{fL}{f\psi_i'(v_i)} \\
-P_i(v_i) &= \frac{\partial}{\partial v_i}\left(T_i(v_i)-v_iP_i(v_i)\right) \\
-P_i(v_i) &= \frac{\partial}{\partial v_i}T_i(v_i)-P_i(v_i)-v_i\frac{\partial}{\partial v_i}P_i(v_i) \\
v_i\frac{\partial}{\partial v_i}P_i(v_i) &= \frac{\partial}{\partial v_i}T_i(v_i) \qquad (13)
\end{aligned}$$

It is important to note that optimization with respect to x, p, and $\psi$ also produces additional boundary conditions that must hold at the values $\underline{v}_i$ and $\overline{v}_i$. These boundary conditions have not been discussed here, however, since it turns out that they are either implications of the above conditions or are satisfied universally.

## 4.5 The Characterization Theorem

We are now in a position to provide a complete characterization of the class of interim efficient mechanisms.

Theorem: $\eta$ = (x,q,t) is an interim efficient mechanism if and only if $\exists\lambda >> 0$ with $\int_{V_i} \lambda_i(v_i) f_i(v_i) dv_i = 1\ \forall i \in N$, such that:

(a) $\forall v \in V$, x(v) and p(v) maximize $\left[\sum_{i=1}^{n} \omega_i(v_i) p_i(v) c_i(p(v)) - K\right] x(v)$

subject to $0 \leq x(v) \leq 1 \quad \forall v \in V$

$0 \leq p_i(v) \leq 1 \quad \forall i \in N, \forall v \in V$

$P_i'(v_i)\ ?\ 0 \quad \forall i \in N, \forall v_i \in V_i$

where
$$\omega_i(v_i) = v_i + \frac{F_i(v_i)}{f_i(v_i)} - \frac{\int_{\underline{v}_i}^{v_i} \lambda_i(y_i) dF_i(y_i)}{f_i(v_i)}$$
$$P_i(w_i) = \int_{V_{-i}} x(w_i, v_{-i}) p_i(w_i, v_{-i}) c_i(p(w_i, v_{-i})) dF_{-i}(v_{-i})$$

(b) $t_i(v) = \int_{\underline{v}_i}^{v_i} w_i P_i'(w_i) dw_i + \tau_i(v)$

where
$$\sum_{i \in N} \tau_i(v) = Kx(v) - \sum_{i \in N} \int_{\underline{v}_i}^{v_i} w_i P_i'(w_i) dw_i, \quad \forall v \in V$$
$$\int_{V_{-i}} \tau_i(v) dF_{-i}(v_{-i}) = 0, \quad \forall i \in N, \forall v_i \in V_i$$

Proof: From equation (7), the first order condition on $t_i(v)$, we have that:

$$\gamma(v) = \lambda_i(v_i) - \frac{\psi_i'(v_i)}{f_i(v_i)} \quad \forall i \in N \tag{14}$$

Note that equality (14) holds regardless of the values $v_j$ for all $j \in N/\{i\}$. Since this is true for all $i \in N$, it follows that $\gamma(v)$ is constant in v. Thus, a rearrangement of the equality gives us:

$$\psi_i'(v_i) = (\lambda_i(v_i) - \gamma) f_i(v_i) \tag{15}$$

Integration with respect to $v_i$ yields:

$$\psi_i(v_i) = \int_{\underline{v}_i}^{v_i} \lambda_i(y_i)\,dF_i(y_i) - \gamma F_i(v_i) + C$$

Applying the boundary conditions $\psi_i(\underline{v}_i) = \psi_i(\bar{v}_i) = 0$ gives us:

$$\psi_i(\underline{v}_i) = \int_{\underline{v}_i}^{\underline{v}_i} \lambda_i(y_i)\,dF_i(y_i) - \gamma F_i(\underline{v}_i) + C = 0$$

$$\Rightarrow \quad C = 0$$

$$\Rightarrow \quad \psi_i(v_i) = \int_{\underline{v}_i}^{v_i} \lambda_i(y_i)\,dF_i(y_i) - \gamma F_i(v_i)$$

$$\psi_i(\bar{v}_i) = \int_{\underline{v}_i}^{\bar{v}_i} \lambda_i(y_i)\,dF_i(y_i) - \gamma F_i(\bar{v}_i) = 0$$

$$\Rightarrow \quad \gamma = \int_{\underline{v}_i}^{\bar{v}_i} \lambda_i(y_i)\,dF_i(y_i) = \bar{\lambda} = 1$$

This gives us:

$$\psi_i(v_i) = \int_{\underline{v}_i}^{v_i} \lambda_i(y_i)\,dF_i(y_i) - F_i(v_i). \tag{16}$$

Substituting (15) and (16) and the result $\gamma = 1$ into the formula for $\omega_i(v_i)$ gives us:

$$
\begin{aligned}
\omega_i(v_i) &= \frac{v_i\lambda_i(v_i)f_i(v_i) - v_i\psi_i'(v_i) - \psi_i(v_i)}{f_i(v_i)} \\
&= \frac{v_i\lambda_i(v_i)f_i(v_i) - v_i(\lambda_i(v_i)f_i(v_i) - f_i(v_i)) - \psi_i(v_i)}{f_i(v_i)} \\
&= \frac{v_i f_i(v_i) - \psi_i(v_i)}{f_i(v_i)} \\
&= v_i - \frac{\psi_i(v_i)}{f_i(v_i)} \\
&= v_i + \frac{F_i(v_i)}{f_i(v_i)} - \frac{\int_{\underline{v}_i}^{v_i} \lambda_i(y_i)\,dF_i(y_i)}{f_i(v_i)}
\end{aligned}
$$

Substituting this formula for $\omega_i(v_i)$ into (6) and applying the boundary condition $\psi_i(\underline{v}_i) = \psi_i(\overline{v}_i) = 0$, the objective function we seek to maximize becomes:

$$
\int_V \left[ x(v) \sum_{i=1}^{n} \left( \omega_i(v_i)p_i(v) - c_i(p(v)) \right) - \overline{\lambda}_i(v_i) - \frac{\psi_i'(v_i)}{f_i(v_i)} + \gamma(v)t_i(v) - \gamma(v)Kx(v) \right] dF(v)
$$

Further substituting conditions (7) and (11), the optimization problem reduces to:

$$
\max_{\eta=(x,p,t)} \int_V \left[ \sum_{i=1}^{n} \omega_i(v_i)p_i(v) - c_i(p(v)) - K \right] x(v)\,dF(v)
$$

which is equivalent to:

$$
\left[ \sum_{i=1}^{n} \omega_i(v_i)p_i(v) - c_i(p(v)) - K \right] x(v) = \max_{\eta=(x,p)} \left[ \sum_{i=1}^{n} \omega_i(v_i)p_i(v) - c_i(p(v)) - K \right] x(v) \quad \forall v \in V
$$

Thus the optimal public good production decision, $x(v)$, and consumption allocation, $p(v)$, must be chosen to maximize this objective function while also satisfying the requirements that $0 \leq x(v) \leq 1$ for all $v \in V$, $0 \leq p_i(v) \leq 1$ for all $i \in N$ and all $v \in V$, and $P_i'(v_i)$ ? 0 for all $i \in N$ and all $\forall v_i \in V_i$. This proves part (a) of the theorem.

Given $x(v)$ and $p(v)$ determined according to the above constrained maximization problem, equations (12) and (13) provide the following first order requirements on the optimal tax shares for allocating the cost of producing the public good:

$$\text{(a)} \quad \sum_{i=1}^{n} t_i(v) = Kx(v)$$

$$\text{(b)} \quad \frac{\partial}{\partial v_i} \int_{V_{-i}} t_i(w_i, v_{-i})\, dF_{-i}(v_{-i}) = v_i \frac{\partial}{\partial v_i} P_i(v_i)$$

Note that (a) is a feasibility and budget balance constraint while (b) is an incentive compatibility constraint. Tax functions that satisfy both of these conditions will be of the form:

$$t_i(v) = \int_{\underline{v}_i}^{v_i} w_i P_i'(w_i)\, dw_i + \tau_i(v) \quad \forall i \in N$$

where $\tau_i(v)$ satisfies the following conditions:

$$\sum_{i \in N} \tau_i(v) = Kx(v) - \sum_{i \in N} \int_{\underline{v}_i}^{v_i} w_i P_i'(w_i)\, dw_i \;, \forall v \in V$$

$$\frac{\partial}{\partial v_i} \int_{V_{-i}} \tau_i(v)\, dF_{-i}(v_{-i}) = 0, \forall i \in N, \forall v_i \in V_i$$

Note that the existence of such $\tau_i(v)$ functions was first shown by d'Aspremont and Gérard-Varet (1979). One example of a family of $\tau_i(v)$ functions that satisfy the described conditions is the following:

$$\tau_i(v)=\frac{K}{n}\bar{x}(v)-X_i(v_i)+\frac{1}{n-1}\sum_{j\neq i} X_j(v_j)-\frac{1}{n-1}\sum_{j\neq i}\int_{\underline{v}_j}^{v_j} w_j P_j'(w_j)\,dw_j$$

where

$$X_i(w_i) = \int_{V_{-i}} x(w_i, v_{-i})\,dF_{-i}(v_{-i}),\ \forall i\in N,\ \forall w_i\in V_i.$$

This proves part (b) of the theorem and hence completes the proof.

Q.E.D.

## 5. Discussion and Interpretation

In this section, we provide some interpretation of the characterization theorem, investigate the issue of second order conditions, and apply the theorem to a particular example of an allocation problem involving a congestible and excludable public good.

### 5.1 Virtual Valuations

The characterization theorem simplifies the optimal production and consumption decisions to the problem of maximizing

$$\left[\sum_{i=1}^{n}\omega_i(v_i)p_i(v)c_i(p(v)) - K\right]x(v) \tag{17}$$

subject to several inequality constraints.

The function $\omega_i(v_i)$ which appears in the characterization theorem is referred to as individual i's *virtual valuation* for the public good given "true" valuation $v_i$. The virtual valuation concept has been extensively employed in similar Bayesian implementation problems (see Myerson 1981, Cornelli 1996, Ledyard and Palfrey 1996). An individual's virtual valuation of the public good is equal to his true valuation of the public good adjusted by a factor that depends on the distribution of these true valuations and on the welfare weights.

For example, consider the nature of these virtual valuations in the benchmark case of "neutral" welfare weights, in which all true valuations are given equal importance in measuring social welfare. In this case, we have $\lambda_i(v_i) = 1$ for all $v_i \in V_i$ which implies $\int_{\underline{v}_i}^{v_i}\lambda_i(y_i)dF_i(y_i) = F_i(v_i)$ and therefore $\omega_i(v_i) = v_i$ for all $i \in N$ and all $v_i \in V_i$. So under neutral welfare weights, an individual's virtual valuation and true valuation are equivalent.

### 5.2 Optimal Consumption

With this notion of virtual valuation in mind, the implications of the characterization theorem for the nature of the optimal public good consumption (or allocation) function, p(v), can be explored by revisiting the first order conditions illustrated in (11). These first order conditions can be reduced to develop the following conditions for optimal consumption in this problem:

$$
\begin{aligned}
p_i(v) = 1 \quad &\Rightarrow \omega_i(v_i)c_i(p) \geq \Delta_i(v,p) \\
p_i(v) \in (0,1) &\Rightarrow \omega_i(v_i)c_i(p) = \Delta_i(v,p) \\
p_i(v) = 0 \quad &\Rightarrow \omega_i(v_i)c_i(p) \leq \Delta_i(v,p)
\end{aligned}
\tag{18}
$$

where

$$\Delta_i(v,p) = \sum_{i=1}^{n} \omega_j(v_j)p_j \frac{f c_j(p)}{f p_i}$$

Note that these conditions could also be uncovered by differentiating the reduced objective function (17) with respect to $p_i(v)$.

These optimal consumption conditions can be interpreted as a determination for each individual whether or not allowing that individual to consume the public good will contribute to the maximization of the objective function (17). In particular, note that inclusion or exclusion of individual i in consumption of the public good will have both a direct and an indirect effect on the value of the summation

$$\sum_{i=1}^{n} \omega_i(v_i)p_i c_i(p)$$

from the objective function. Allowing individual i to consume the public good, by setting $p_i(v)=1$, directly increases this summation by the amount $\omega_i(v_i)c_i(p)$. On the other hand, inclusion of individual i will also have an indirect effect by changing the value of $c_j(p)$ for all individuals and therefore the summation will also be adjusted by the amount $\Delta_i(v,p)$. Thus, allowing individual i to consume the public good will increase the value of the objective function (17) if and only if $\omega_i(v_i)c_i(p) + \Delta_i(v_{-i},p) \geq 0$ or, alternatively, $\omega_i(v_i)c_i(p) \geq -\Delta_i(v_{-i},p)$, which is precisely the first order inequality constraint given in (18). Note that when congestion creates a negative externality, $\Delta_i(v,p)$ will be positive and the first order condition says that individual i can be allowed to consume the public good if and only if the benefit to individual i outweighs the congestion cost to all individuals (including possibly herself) who consume the public good.

### 5.3 Optimal Production

Now, let us similarly explore the implications of the characterization theorem for the nature of the optimal public good production function, x(v). First note that the objective function (17) is simply x(v) multiplied by the quantity

$$\sum_{i=1}^{n} \omega_i(v_i)p_i(v)c_i(p(v)) - K$$

which is independent of x(v). Therefore, the objective function will be maximized by setting x(v) = 1 (i.e., producing the maximum permitted level of the public good) whenever we have:

$$\sum_{i=1}^{n} \omega_i(v_i)p_i c_i(p) \geq K$$

Similarly, the objective function will be maximized by setting x(v) = 0 (i.e., producing the minimum permitted level of the public good) whenever we have:

$$\sum_{i=1}^{n} \omega_i(v_i)p_i c_i(p) \leq K$$

The nature of the objective function allows us to effectively ignore solutions involving $x(v) \in (0,1)$ because this can only be optimal when we have:

$$\sum_{i=1}^{n} \omega_i(v_i)p_i c_i(p) = K$$

in which case the objective function always attains a value of zero, regardless of the value of $x(v)$ and, in particular, $x(v) = 1$ and $x(v) = 0$ will both maximize the objective function in such situations. Therefore, it is accurate to say that it is always optimal to either set $x(v) = 1$ or to set $x(v) = 0$.

Thus, we can effectively write the optimal production condition as:

$$\begin{aligned} x(v) = 1 &\Leftrightarrow \sum_{i=1}^{n} \omega_i(v_i)p_i c_i(p) \geq K \\ x(v) = 0 &\Leftrightarrow \sum_{i=1}^{n} \omega_i(v_i)p_i c_i(p) < K \end{aligned} \tag{19}$$

To further interpret this optimal production condition, recall that individual i's utility from consuming a proportion $p_i$ of one unit of the public good is given by $v_i p_i c_i(p)$. Thus, the value $\omega_i(v_i)p_i c_i(p)$, which appears in (19) as well as the objective function (17), can be seen as individual i's "virtual" utility from consuming a proportion $p_i$ of one unit of the public good, in which individual i's virtual valuation replaces her true valuation in the utility function. Thus, condition (19) can be interpreted to say that it is optimal to produce the public good whenever the sum of individual virtual utilities from consuming the allocated proportions of the public good is greater than or equal to the unit cost of producing the public good.

## 5.4 Second Order Conditions

The analysis so far has not explicitly addressed whether the second order conditions for a maximum are satisfied. These second order conditions are satisfied as long as $\pi_i(v_i)$ is convex in $v_i$, which is true whenever:

$$\frac{\partial \pi_i(v_i)}{\partial v_i} = P_i(v_i)$$

is weakly increasing in $v_i$ or, alternatively, $P_i'(v_i) \geq 0$. Thus, the second order conditions will be satisfied if the interim expected public good consumption for individual i (adjusted for congestion) is not decreasing in individual i's public good valuation. This is a sensible condition, since we would expect that a higher public good valuation would not result in a lower level of public good consumption.

In the characterization theorem, we have explicitly applied the constraint $P_i'(v_i) \geq 0$ which, as described in section 4.2, is required for incentive compatibility. It is instructive, however, to investigate whether there are additional conditions which ensure that maximization of the *relaxed program*, the optimization program in the characterization theorem <u>without</u> the $P_i'(v_i) \geq 0$ condition, also satisfies this second order constraint. With this in mind, consider the following assumption:

<u>Assumption 1:</u> For all $i \in N$, $\omega_i(v_i)$ is a strictly increasing function of $v_i$.

This assumption places a restriction on the welfare weighting functions $(\lambda_i(v_i) \text{ for } i = 1, 2, \ldots, n)$ and on the distribution functions for individual values $(F_i(v_i) \text{ for } i = 1, 2, \ldots, n)$. We have already shown in section 5.1 that this assumption is satisfied for the case of neutral welfare weights (where $\omega_i(v_i) = v_i$), and it will also be satisfied when $\lambda_i(v_i)$ and $F_i(v_i)$ are given by other well-known distributions. Situations in which Assumption 1 is satisfied are commonly referred to as the "regular" case (Myerson 1981).

We now demonstrate that Assumption 1 is sufficient to ensure that optimization of the relaxed program produces a $P_i(v_i)$ that is weakly increasing in $v_i$, and therefore that the general program is also optimized.

<u>Proposition 1:</u> Suppose x(v) and p(v) are optimal in the relaxed program. If Assumption 1 holds, then $P_i(v_i) = \int_{V_{-i}} x(v) p_i(v) c_i(p(v)) dF_{-i}(v_{-i})$ is weakly increasing in $v_i$.

<u>Proof:</u> First note from the formula for $P_i(v_i)$ that whenever x(v) and the product $p_i(v) c_i(p(v))$ are both weakly increasing in $v_i$, $P_i(v_i)$ will also be weakly increasing in $v_i$. We therefore proceed by proving in separate claims that x(v) and $p_i(v) c_i(p(v))$ are each weakly increasing in $v_i$ under Assumption 1.

<u>Claim 1:</u> If Assumption 1 holds, then $\forall i \in N$ and $\forall v \in V$, x(v) is weakly increasing in $v_i$.

<u>Proof:</u> Suppose, on the contrary, that for some $i \in N$ and some $v \in V$ with optimal consumption vector p(v) and optimal public good production x(v), there exists $v' \in V$ where $v'_i > v_i$ and $v'_j = v_j$ $\forall j \neq i$, with optimal consumption vector $p(v')$ and optimal public good production $x(v')$, such that $x(v') < x(v)$.

$$x \in [0,1] \text{ and } x(v') < x(v) \Rightarrow x(v) \in (0,1] \text{ and } x(v') \in [0,1)$$

$$x(v) \in (0,1] \Rightarrow \sum_{i=1}^{n} \omega_i(v_i) p_i(v) c_i(p(v)) - K \geq 0$$

$$x(v') \in [0,1) \Rightarrow \sum_{i=1}^{n} \omega_i(v_i') p_i(v') c_i(p(v')) - K \leq 0$$

$$\Rightarrow \sum_{i=1}^{n} \omega_i(v_i') p_i(v') c_i(p(v')) \leq \sum_{i=1}^{n} \omega_i(v_i) p_i(v) c_i(p(v)) \tag{20}$$

The optimality of $p(v')$ under valuation vector $v'$ requires that

$$\sum_{i=1}^{n} \omega_i(v_i') p_i(v') c_i(p(v')) \geq \sum_{i=1}^{n} \omega_i(v_i') p_i(v) c_i(p(v)) \tag{21}$$

By Assumption 1, however, we have $\omega_i(v_i') > \omega_i(v_i)$ since $v_i' > v_i$. For all j≠i, we have $v_j' = v_j$ and thus $\omega_j(v_j') = \omega_j(v_j)$. Therefore, we have

$$\sum_{i=1}^{n} \omega_i(v_i') p_i(v) c_i(p(v)) > \sum_{i=1}^{n} \omega_i(v_i) p_i(v) c_i(p(v)).$$

Combining this with (20) gives us

$$\sum_{i=1}^{n} \omega_i(v_i') p_i(v) c_i(p(v)) > \sum_{i=1}^{n} \omega_i(v_i') p_i(v') c_i(p(v'))$$

which contradicts (21). Therefore, it must be the case that x(v) is weakly increasing in $v_i$.

Note that if x(v) = 0 then the total product $x(v) p_i(v) c_i(p(v))$ is also zero and therefore must be weakly increasing in $v_i$ at that point (since $x(v) p_i(v) c_i(p(v))$ can not be negative).

Therefore, in the next claim, we only need to demonstrate validity at points where $x(v) > 0$.

<u>Claim 2:</u> If Assumption 1 holds, then $\forall i \in N$ and $\forall v \in V$ with $x(v) > 0$, $p_i(v)c_i(p(v))$ is weakly increasing in $v_i$.

<u>Proof:</u> Suppose on the contrary that $\exists i \in N$ and $\exists v, v' \in V$ with $v'_i > v_i$ and $v'_j = v_j \; \forall j \neq i$, such that $x(v) > 0$ and $p_i(v)c_i(p(v)) > p_i(v')c_i(p(v'))$ where x and p are optimal production and consumption functions respectively. Note that, since $v'_i > v_i$, we have $x(v'_i) > 0$ by Claim 1.

The optimality of p(v) implies:

$$\omega_i(v_i)p_i(v)c_i(p(v)) + \underset{j?k}{} \omega_j(v_j)p_j(v)c_j(p(v)) \geq \omega_i(v_i)p_i(v')c_i(p(v')) + \underset{j?k}{} \omega_j(v_j)p_j(v')c_j(p(v'))$$

Similarly, the optimality of $p_i(v')$ implies:

$$\omega_i(v'_i)p_i(v')c_i(p(v')) + \underset{j?k}{} \omega_j(v_j)p_j(v')c_j(p(v')) \geq \omega_i(v'_i)p_i(v)c_i(p(v)) + \underset{j?k}{} \omega_j(v_j)p_j(v)c_j(p(v))$$

Combining these two inequalities gives us:

$$\begin{aligned}
&\omega_i(v_i)p_i(v)c_i(p(v)) + \underset{j?k}{} \omega_j(v_j)p_j(v)c_j(p(v)) \\
&\quad \geq \omega_i(v_i)p_i(v')c_i(p(v')) + \omega_i(v'_i)p_i(v)c_i(p(v)) + \underset{j?k}{} \omega_j(v_j)p_j(v)c_j(p(v)) - \omega_i(v'_i)p_i(v')c_i(p(v'))
\end{aligned}$$

$$\Rightarrow \omega_i(v_i)p_i(v)c_i(p(v)) \geq \omega_i(v_i)p_i(v')c_i(p(v')) + \omega_i(v'_i)p_i(v)c_i(p(v)) - \omega_i(v'_i)p_i(v')c_i(p(v'))$$

$$\Rightarrow \omega_i(v_i)[p_i(v)c_i(p(v)) - p_i(v')c_i(p(v'))] \geq \omega_i(v'_i)[p_i(v)c_i(p(v)) - p_i(v')c_i(p(v'))]$$

By Assumption 1, we have $\omega_i(v_i') > \omega_i(v_i)$ since $v_i' > v_i$, which means that this last inequality holds if and only if $p_i(v)c_i(p(v)) \leq p_i(v')c_i(p(v'))$, which contradicts our initial assumption. Therefore, it must be the case that $p_i(v)c_i(p(v))$ is weakly increasing in $v_i$.

Q.E.D.

### 5.5 Example: A Two Person Allocation Problem

Let N = {1,2} and suppose that we have neutral welfare weights $(\text{i.e.}, \lambda_i(v_i)=1, \quad \forall i \in N, \forall v_i \in V_i)$, so that our optimization problem selects the first best solution. Further suppose that the congestion functions are given by:

$$c_1(p) = 1 - \tfrac{1}{2}p_2$$

$$c_2(p) = 1 - \tfrac{1}{4}p_1$$

In this case, the optimal consumption conditions illustrated by the formulae in (18) become:

$$p_1(v) = 1 \quad \Rightarrow v_1(1 - \tfrac{1}{2}p_2)\ ?\ \tfrac{1}{4}v_2p_2$$

$$p_1(v) \in (0,1) \Rightarrow v_1(1 - \tfrac{1}{2}p_2) = \tfrac{1}{4}v_2p_2$$

$$p_1(v) = 0 \quad \Rightarrow v_1(1 - \tfrac{1}{2}p_2) \leq \tfrac{1}{4}v_2p_2$$

$$p_2(v) = 1 \quad \Rightarrow v_2(1 - \tfrac{1}{4}p_1)\ ?\ \tfrac{1}{2}v_1p_1$$

$$p_2(v) \in (0,1) \Rightarrow v_2(1 - \tfrac{1}{4}p_1) = \tfrac{1}{2}v_1p_1$$

$$p_2(v) = 0 \quad \Rightarrow v_2(1 - \tfrac{1}{4}p_1) \leq \tfrac{1}{2}v_1p_1$$

We can quickly eliminate the possibility that $p_1(v) \in (0,1)$ and $p_2(v) \in (0,1)$, since there does not exist a combination of permissible $v_1$, $v_2$, $p_1$, and $p_2$ that allow both necessary equalities to be satisfied simultaneously. Moreover, we can ignore the possibility that $p_1(v)$ is 0 or 1 while $p_2(v) \in (0,1)$, since there is only one value of $v_2$ for which the necessary equality will hold for any given $v_1$ in this case, and drawing this one $v_2$ is a zero probability event. Similarly, we can ignore the possibility that $p_2(v)$ is 0 or 1 while $p_1(v) \in (0,1)$, since there is only one value of $v_1$ for which the necessary equality will hold for any given $v_2$ in this case.

Thus, we can effectively rewrite the optimal consumption conditions as:

$$p_1(v) = 1 \Leftrightarrow v_1\left(1 - \tfrac{1}{2}p_2\right) ? \tfrac{1}{4}v_2 p_2$$
$$p_1(v) = 0 \Leftrightarrow v_1\left(1 - \tfrac{1}{2}p_2\right) < \tfrac{1}{4}v_2 p_2$$

$$p_2(v) = 1 \Leftrightarrow v_2\left(1 - \tfrac{1}{4}p_1\right) ? \tfrac{1}{2}v_1 p_1$$
$$p_2(v) = 0 \Leftrightarrow v_2\left(1 - \tfrac{1}{4}p_1\right) < \tfrac{1}{2}v_1 p_1$$

This give us:

$$p = (1,1) \Leftrightarrow 2v_1 \geq v_2 \geq \tfrac{2}{3}v_1$$
$$p = (1,0) \Leftrightarrow v_1 \geq 0 \text{ and } v_2 < \tfrac{2}{3}v_1$$
$$p = (0,1) \Leftrightarrow v_2 \geq 0 \text{ and } v_1 < \tfrac{1}{2}v_2$$

The optimal production decision is given by:

For $p = (1,1)$, $x(v) = 1 \Leftrightarrow \frac{1}{2}v_1 + \frac{3}{4}v_2 \geq K$

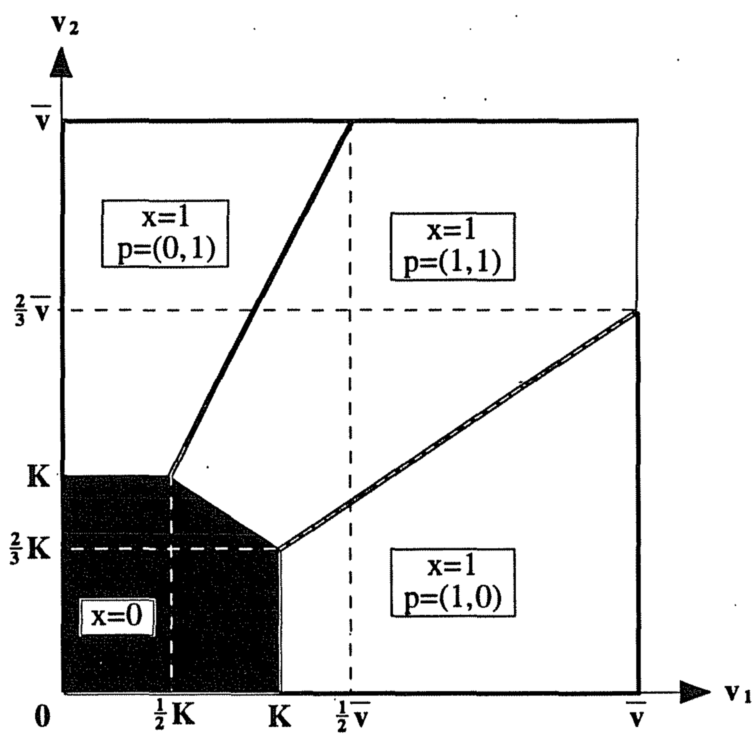
For $p = (1,0)$, $x(v) = 1 \Leftrightarrow v_1 \geq K$

For $p = (0,1)$, $x(v) = 1 \Leftrightarrow v_2 \geq K$

Thus, in Figure 1 the optimal production and consumption decisions are illustrated for the case when $V_1 = V_2 = [0, \bar{v}]$ and $0 < K < \bar{v}$.

Figure 1

Optimal Production and Consumption in the Two Person Problem




Thus, the probability that the public good is produced is increasing in both valuations, $v_1$ and $v_2$, while the probability that individual i will consume the public good is increasing in her own valuation $v_i$. Moreover, as $v_i$ increases it also becomes more likely that individual i will be the <u>only</u> consumer of the public good.

It is also clear from Figure 1 that this solution satisfies the second order condition requiring $P_i'(v_i) \geq 0$. The figure demonstrates that, as $v_i$ increases, the expected value of x increases, the expected value of $p_i$ increases, and, since the expected value of $p_j$ decreases for $j \neq i$, the expected value of $c_i(p)$ increases. Thus each component of

$$P_i(v_i) = {}_{V_{-i}} x(v)p_i(v)c_i(p(v))dF_{-i}(v_{-i})$$

is increasing in $v_i$, thus we have $P_i'(v_i) \geq 0$ and the second order condition is satisfied.

Figure 1 also illustrates that, in the optimal solution, it is more likely that individual 1 consumes the public good than that individual 2 consumes the public good. This is because we have

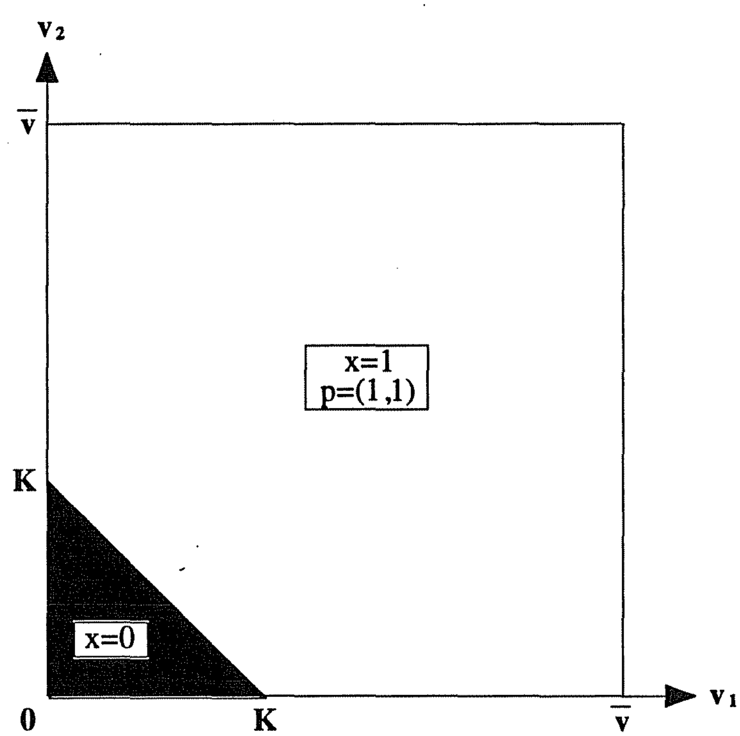
$$-\frac{\partial c_2(p)}{\partial p_1} = \frac{1}{4} < \frac{1}{2} = -\frac{\partial c_1(p)}{\partial p_2}$$

and thus individual 1's consumption has a smaller negative effect on individual 2's consumption utility than individual 2's consumption has on individual 1's consumption utility.

To demonstrate the effect of the presence of congestion in this example, compare Figure 1 with Figure 2, in which the optimal production and consumption decisions are illustrated for the no congestion case. Without congestion, the public good is produced whenever we have $v_1+v_2 \geq K$ and neither individual is ever excluded from consumption (the condition for exclusion of individual i in the no congestion case is $v_i < 0$). Thus, congestion reduces the probability that the public good is produced and, even when the public good is produced, congestion reduces each individual's probability of consumption. Note that the no congestion case will be discussed in more detail and generality in section 6.1.

Figure 2

Optimal Production and Consumption in the Two Person Problem without Congestion

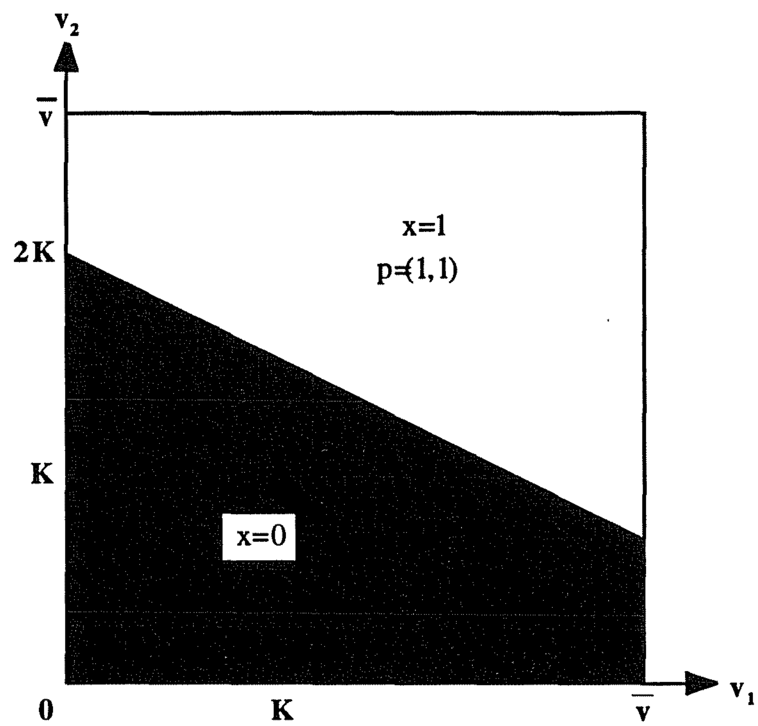



To now demonstrate the effect of excludability in this example, compare Figure 1 with Figure 3, in which the optimal production and consumption decisions are illustrated for the congestion but no exclusion case, for which the constraint p = (1,1) is imposed.

Without exclusion, the public good is produced whenever we have $\frac{1}{2}v_1 + \frac{1}{4}v_2 \geq K$ which, as illustrate in Figure 3, is a much more stringent requirement than in the excludable case. Thus, the presence of excludability in this example increases the probability that the public good is produced. Note, however, that there are some valuation pairs in the non-excludable case for which the public good is produced and is consumed by both individuals while in the excludable case, for the same valuation pair, only one individual consumes the public good. The no exclusion case will be discussed in more detail and generality in section 6.2.

<u>Figure 3</u>

Optimal Production and Consumption in the Two Person Problem without Exclusion



## 6. Special Cases

In this section we discuss the implications of our general characterization theorem for several special classes of allocation problems. These specialized problems include situations of no congestion, no exclusion, as well as cases in which certain conditions are placed on the individual congestion functions.

### 6.1 No Congestion

Suppose that the public good is excludable but that consumption of the good creates no congestion. In this case we have that:

$$c_i(p) = 1,\ \forall i \in N,\ \forall p \in [0,1]^n$$
$$\frac{\partial c_j(p)}{\partial p_i} = 0,\ \forall i,j \in N,\ \forall p \in [0,1]^n$$

Therefore, the optimal exclusion condition (18) becomes:

$$\begin{aligned} p_i(v) = 1 \quad & \Rightarrow \omega_i(v_i) \geq 0 \\ p_i(v) \in (0,1) & \Rightarrow \omega_i(v_i) = 0 \\ p_i(v) = 0 \quad & \Rightarrow \omega_i(v_i) \leq 0 \end{aligned}$$

Thus, in the optimal solution, the only individuals who are excluded from consumption of the public good are those who have a negative (or zero) virtual valuation for the public good. Let $\omega_+ = \{i \in N \mid \omega_i(v_i) > 0\}$ be the set of individuals who are included in consumption of the public good. With this in mind, the optimal production condition (19) in this situation becomes:

$$x(v) = 1 \Leftrightarrow \sum_{i \in \omega_+} \omega_i(v_i) \geq K$$
$$x(v) = 0 \Leftrightarrow \sum_{i \in \omega_+} \omega_i(v_i) < K$$

Thus, in the optimal solution, the public good is produced if and only if the sum of all positive virtual valuations is greater than or equal to the cost, K, of producing the public good.

Note that in deriving optimal production conditions for this special case, we can effectively ignore individuals for whom $\omega_i(v_i) = 0$. Although such individuals may consume a positive quantity of the public good in an optimal solution, their inclusion or exclusion does not affect the value of the objective function (17) in this no congestion case, since $\omega_i(v_i)p_i(v)$ will always be zero for such individuals. Therefore, whether or not such individuals are included in consumption of the public good (and, if so, at what quantity) has no impact on the decision whether or not to produce the public good.

## 6.2 No Exclusion

Now suppose that the public good is congestible but it is not excludable, so that all individuals must consume the entire quantity of the public good if it is produced. In this case, we have $p_i = 1$ for all $i \in N$, and there is no first order condition on p, and therefore no optimal exclusion condition. It can be shown that all other conditions of the theorem hold with the substitution $p_i = 1$ for all $i \in N$. Thus, the optimal production decision is given by:

$$x(v) = 1 \Leftrightarrow \sum_{i=1}^{n} \omega_i(v_i) c_i(1,1,\ldots,1) \geq K$$
$$x(v) = 0 \Leftrightarrow \sum_{i=1}^{n} \omega_i(v_i) c_i(1,1,\ldots,1) < K$$

Hence, the optimal solution is to produce the public good if and only if the sum of all virtual valuations, adjusted for congestion, is greater than or equal to the cost, K, of producing the public good. To compare this to the excludable case, consider the following proposition which tells us that it is always less likely that the public good will be produced in the non-excludable case than when exclusion is permitted.

<u>Proposition 2</u>: For any vector of value reports $v = \{v_1, v_2, \ldots, v_n\}$, if it is optimal to produce the public good under no exclusion, then it will also be optimal to produce the public good when exclusion is permitted. The converse is not true, however.

<u>Proof:</u> Suppose that x=1, that is, producing the public good, is optimal in the no exclusion case. Thus, we have:

$$\sum_{i=1}^{n} \omega_i(v_i) c_i(1,1,\ldots,1) \geq K. \tag{22}$$

If the consumption probability vector $p = (1,1,\ldots,1)$ is optimal even when exclusion is permitted, then the proof is complete, because condition (22) would be sufficient for x=1 to also be optimal in the excludable case. Thus, suppose instead that $p = (1,1,\ldots,1)$ is <u>not</u> optimal when exclusion is permitted. Note, however, that (22) still holds in this case, so having $p = (1,1,\ldots,1)$ non-optimal means that there must be some other optimal $p' \in [0,1]^n$ for which

$$\sum_{i=1}^{n} \omega_i(v_i)p'_i(v)c_i(p(v)) > \sum_{i=1}^{n} \omega_i(v_i)c_i(1,1,\ldots,1)\,.$$

Combining this inequality with (22) and comparing to the optimal production condition (19) illustrates that it is still optimal to have x=1 in this case. Thus, whenever the public good is produced under no exclusion, it would also be produced if exclusion were permitted.

To prove the second part of the proposition, we need only refer to the example in section 5.5, in which there were valuation vectors for which production of the public good was optimal with excludability, but not optimal without excludability.

Q.E.D.

### 6.3 Camaraderie or Complementarity

Suppose that there is some individual i whose consumption of the public good generates a positive (or non-negative) externality. That is, individual i's consumption provides a value of camaraderie or complementarity to the other individuals who consume the public good. Mathematically speaking, this means that:

$$\frac{\partial c_j(p)}{\partial p_i} \geq 0, \; \forall j \in N \qquad (23)$$

An individual for whom condition (23) applies can be understood to be someone with whom all other individuals wish to associate in terms of shared consumption of the public good. Before evaluating the implications of this condition for the optimal consumption vector, consider the following assumption.

<u>Assumption 2:</u> $c_i(p) \geq 0 \quad \forall i \in N \ \forall p \in [0,1]^n$

This assumption says that congestion effects can not change the sign of an individual's utility from consuming the public good. In other words, if an individual has a positive valuation for the public good, no amount of congestion will cause this individual to have negative consumption utility. Note that Assumption 3 will always be satisfied for an individual with positive valuation when there is free disposal of the public good. That is, if individual i with valuation $v_i \geq 0$ can not be compelled to consume the public good, it must be the case that $c_i(p) \geq 0$ for all $p \in [0,1]^n$.

When this assumption is satisfied, we have the following result.

<u>Proposition:</u> Suppose Assumption 2 holds and that there exists some individual $i \in N$ for whom $\omega_i(v_i) \geq 0$ and condition (23) is satisfied. Then it will always be optimal to have $p_i$=1 in the consumption probability vector p.

<u>Proof:</u> Recall that it will be optimal to have $p_i$=1 in the consumption probability vector p whenever the following inequality is satisfied:

$$\omega_i(v_i)c_i(p) ? -\sum_{j=1}^{n} \omega_j(v_j)p_j \frac{\partial c_j(p)}{\partial p_i}.$$

For any individual i for whom $\dot{\omega}_i(v_i) \geq 0$ we will have $\omega_i(v_i)\ c_i(p) \geq 0$ because $c_i(p) \geq 0$ by Assumption 1. Therefore, for such an individual, the left-hand side of the above inequality is either zero or positive. Note that Assumption 1 guarantees that $\omega_j(v_j)p_j$ will be non-negative for all j in the optimal solution (since the optimal solution maximizes the sum of $\omega_j(v_j)p_j\ c_j(p)$ across all individuals j). Therefore, if individual i also satisfies (23), the right-hand side of this inequality will be either zero or negative, and the inequality will therefore always be satisfied. Thus, it will always be optimal to have $p_i$=1 in the consumption probability vector p.

Q.E.D.

This result implies that whenever a congestible and excludable public good must be allocated under Assumption 2, it is optimal to first include in consumption all individuals with positive virtual valuations who provide camaraderie benefits to others, and to then determine how to allocate consumption among the remaining individuals. Note that it might be the case that <u>all</u> individuals provide camaraderie benefits to others. Such situations may arise, for example, when the excludable public good is a telephone network, where individual utility is increasing in the number of other individuals who are connected. In such cases, it is clear that no individual with positive virtual valuation will be excluded from consumption.

### 6.4 Identical and Anonymous Congestion Effects

Suppose that all individuals experience identical congestion effects. In particular, suppose that the following assumption holds.

<u>Assumption 3:</u> $c_i(p) = c_j(p) \equiv c(p) \quad \forall i,j \in N, \forall p \in [0,1]^n$.

Now also suppose that congestion effects are anonymous or, in other words, that an individual's consumption utility is affected equally by the consumption of all other individuals. In particular, suppose that the following assumption holds.

<u>Assumption 4:</u> $\dfrac{\partial c_i(p)}{\partial p_j} = \dfrac{\partial c_i(p)}{\partial p_k} \equiv \dfrac{\partial c_i(p)}{\partial p_x} \quad \forall i,j,k \in N, \forall p \in [0,1]^n$.

Note that combining these two assumptions gives us

$$\frac{\partial c_i(p)}{\partial p_k} = \frac{\partial c_j(p)}{\partial p_k} = \frac{\partial c_j(p)}{\partial p_l} \equiv \frac{\partial c(p)}{\partial p_x} \quad \forall i,j,k,l \in N$$

Thus, under Assumptions 3 and 4 (and Assumption 1 from section 5.4), the optimal consumption conditions become:

$$\begin{aligned}
p_i(v) = 1 \quad &\Rightarrow \omega_i(v_i) \geq -\frac{1}{c(p)}\frac{\partial c(p)}{\partial p_x}\sum_{j \neq i}\omega_j(v_j)p_j \\
p_i(v) \in (0,1) &\Rightarrow \omega_i(v_i) = -\frac{1}{c(p)}\frac{\partial c(p)}{\partial p_x}\sum_{j \neq i}\omega_j(v_j)p_j \\
p_i(v) = 0 \quad &\Rightarrow \omega_i(v_i) \leq -\frac{1}{c(p)}\frac{\partial c(p)}{\partial p_x}\sum_{j \neq i}\omega_j(v_j)p_j
\end{aligned} \tag{24}$$

Under these assumptions, we have a simple characterization of the optimal consumption vector. In particular, the optimal consumption vector will have all individuals with a public good valuation above some threshold consuming the public good (assuming it is produced), and all individuals with a public good valuation below some threshold not consuming the public good. This result is formalized in the following proposition.

<u>Proposition 3:</u> Under Assumption 1, Assumption 3, and Assumption 4, suppose that individual values are all distinct and are ordered such that $v_1 > v_2 > \cdots > v_n$. For each $k \in \{0,1,\ldots,n\}$, let $p^k = \{p_1^k, p_2^k, \ldots, p_n^k\}$ identify a family of consumption vectors given by $p_i^k = 1$ for $i < k$, $p_i^k \in (0,1]$ for $i = k$, and $p_i^k = 0$ for $i > k$. Then for exactly one $k \in \{0,1,2,\ldots,n\}$, $p^k$ will characterize all optimal consumption vectors.

<u>Proof:</u> We prove this proposition by proving three claims about the nature of the optimal consumption vector which combine to say that the optimal consumption vector must be of the form $p^k$ described in the proposition.

<u>Claim 1:</u> Any optimal consumption vector $p^*(v)$ will have at most one $i \in N$ such that $p_i^*(v) \in (0,1)$.

To prove this claim, suppose on the contrary that $p^*(v)$ is optimal with $p_i^*(v) \in (0,1)$ and $p_j^*(v) \in (0,1)$ where $i < j$ and therefore $v_i > v_j$ and, by Assumption 1, $\omega_i(v_i) > \omega_j(v_j)$. By the optimal consumption condition (24), we have

$$\omega_i(v_i) = -\frac{1}{c(p^*)}\frac{\partial c(p^*)}{\partial p_x}\sum_{k=1}^{n}\omega_k(v_k)p_k^*$$

$$\omega_j(v_j) = -\frac{1}{c(p^*)}\frac{\partial c(p^*)}{\partial p_x}\sum_{k=1}^{n}\omega_k(v_k)p_k^*$$

$$? \quad \omega_i(v_i) = \omega_j(v_j)$$

However, this last equality cannot hold, since we know $\omega_i(v_i) > \omega_j(v_j)$. Therefore, we cannot have an optimal consumption vector $p^*(v)$ with $p_i^*(v) \in (0,1)$ $p_j^*(v) \in (0,1)$ where i≠j.

Claim 2: If, in any optimal consumption vector $p^*(v)$, we have $p_i^*(v) \in (0,1]$, then $p_j^*(v) = 1$ for all j<i and $p_j^*(v) = 0$ for all j>i.

To prove this claim, suppose on the contrary that $p^*(v)$ is optimal with $p_i^*(v) \in (0,1]$ and $p_j^*(v) \in [0,1)$ where j<i and therefore $v_i < v_j$ and, by Assumption 1, $\omega_i(v_i) < \omega_j(v_j)$. By the optimal consumption condition (24), we have

$$\omega_i(v_i)c_i(p^*) ? - \frac{fc(p^*)}{fp_x} \sum_{k=1}^{n} \omega_k(v_k)p_k^*$$

$$\omega_j(v_j)c_j(p^*) \leq - \frac{fc(p^*)}{fp_x} \sum_{k=1}^{n} \omega_k(v_k)p_k^*$$

$$? \quad \omega_i(v_i)c_i(p^*) ? \ \omega_j(v_j)c_j(p^*)$$

However, this last inequality cannot hold, since $\omega_i(v_i) < \omega_j(v_j)$ and $c_i(p^*) = c_j(p^*)$. Therefore, we must have $p_j^*(v) = 1$.

Claim 3: If, in any optimal consumption vector $p^*(v)$, we have $p_i^*(v) \in [0,1)$, then $p_j^*(v) = 0$ for all j>i.

To prove this claim, suppose on the contrary that $p^*(v)$ is optimal with $p_i^*(v) \in [0,1)$ and $p_j^*(v) \in (0,1]$ where j>i and therefore $v_i > v_j$ and, by Assumption 1, $\omega_i(v_i) > \omega_j(v_j)$. By the optimal consumption condition (24), we have

$$\omega_i(v_i)c_i(p^*) \leq -\frac{\partial c(p^*)}{\partial p_x}\sum_{k=1}^{n} \omega_k(v_k)p_k^*$$
$$\omega_j(v_j)c_j(p^*) \geq -\frac{\partial c(p^*)}{\partial p_x}\sum_{k=1}^{n} \omega_k(v_k)p_k^*$$
$$\Rightarrow \quad \omega_i(v_i)c_i(p^*) \leq \omega_j(v_j)c_j(p^*)$$

However, this last inequality cannot hold, since $\omega_i(v_i) > \omega_j(v_j)$ and $c_i(p^*) = c_j(p^*)$. Therefore, we must have $p_j^*(v) = 0$.

Putting claims 1 through 3 together, we have that any optimal consumption vector will be characterized by the description of the family of vectors $p^k$ for exactly one $k \in \{0,1,\ldots,n\}$.

Q.E.D.

## 7. Conclusions

In this paper, we explore the issue of efficient allocation of an excludable and congestible public good. In particular, we focus on an economic environment in which the public good is produced at constant returns to scale up to a maximum feasible level, and in which individuals have independent private valuations for the public good and congestion functions which adjust their consumption utility based on the consumption of others. An allocation rule in this environment consists of three elements: (1) a decision rule that determines the level of the public good produced, (2) a condition which determines which individuals will consume the public good and which will be excluded from consumption, and (3) a set of tax functions which distribute the cost of producing the public good across individuals.

We have fully characterized the set of interim efficient allocation rules in this asymmetric information environment using a Bayesian implementation approach. We find that the description of optimal allocation rules relies heavily upon the use of the concept of virtual valuation, which is a function of the true public good valuation, the probability distribution of valuations, and the welfare weighting function for each particular individual. The optimal exclusion condition in this environment is dependent not only on the virtual valuations, but also on the individual congestion functions and the derivatives of these congestion functions with respect to the consumption of each other individual. In general, optimal exclusion requires that an individual be excluded from consumption if and only if his inclusion would lower the sum of included virtual valuations adjusted for congestion. In other words, if the negative congestion effect an individual's consumption creates is greater than the positive benefit that this individual gets from consuming the public good, then that individual must be excluded. The optimal public good production solution is then to produce the maximum feasible level of the public good whenever the sum of the included virtual valuations adjusted for congestion is greater than the cost of production.

We demonstrate that the conclusions of this analysis can be adapted to characterize the set of interim efficient mechanisms in environments without exclusion or congestion, and find that the resulting characterization is analogous to the previous work of Ledyard and Palfrey (1996) and Cornelli (1996). We also demonstrate that when congestion effects are negative, it is less likely that the public good will be produced than in the absence of congestion, and moreover, even when the public good is produced, it is less likely in the presence of congestion that each individual will be permitted to consume the public good. When an individual's consumption produces positive congestion effects, on the other hand, we demonstrate that such an individual will always be included in consumption of the public good whenever his virtual valuation is positive.

Lastly, we applied our characterization to the particular case in which individuals experience identical and anonymous congestion effects. We find in this case, that the optimal set of consumers of the public good is all individuals whose virtual valuations are greater than or equal to some particular threshold. This is analogous to the result of Cornelli (1996) who found, in the absence of congestion, that the optimal set of consumers was all individuals whose virtual valuations were greater than or equal to zero. The result under congestion is nonetheless significantly different, however, because the threshold first depends on the individual congestion functions but, more importantly, it also depends on the actual realization of all individual valuations (i.e., not just the prior probability distribution of valuations). Thus, in the absence of congestion, an individual with a particular realized virtual valuation will always know, without any information from other individuals, whether or not she will be consuming the public good (assuming it is produced). In the presence of negative congestion effects, however, even an individual with a very high (or possibly very low) virtual valuation may have to wait until all valuations are revealed before knowing whether or not she will be included in or excluded from consumption.

The most important extension of the current research would be to further explore the nature of the non-linear tax functions described in the characterization theorem. In particular, it would be helpful to determine if there exists an optimal indirect mechanism in this economic environment in which, instead of individuals reporting a valuation for the public good, they choose to pay a particular tax, which may or may not be conditional on actual production of the public good. It is conjectured that, in the special case of identical congestion functions that we examined, such an indirect mechanism would be analogous to the indirect mechanism for excludable but non-congestible public goods described by Cornelli (1996). In the non-congestible case, Cornelli found that the optimal direct mechanism involved providing the public good, conditional on production, to any individual who offers to pay a tax above a particular predetermined threshold. In the congestible case, we might expect a similar result; however, the particular threshold would not be determined until after the individual reports are realized.

## REFERENCES

Arnott, R., A. de Palma, and R. Lindsey. 1992. "Route Choice with Heterogenous Drivers and Group-Specific Congestion Costs." *Regional Science and Urban Economics* 22:71-102.

Bennett, E., and D. Conn. "The Group Incentive Properties of Mechanism for the Provisision of Public Goods." *Public Choice* 24:95-102

Berglas, E., and D. Pines. 1981. "Clubs, Local Public Goods and Transportation Models: A Synthesis." *Journal of Public Economics* 15:141-162.

Clark, C. 1980. "Restricted Access to Common-Property Fishery Resources: A Game Theoretic Analysis." In *Dynamic Optimization and Mathematical Economics*, edited by P.-T. Liu. New York: Plenum.

Clarke, E.H. 1971. "Multipart Pricing of Public Goods." *Public Choice* 8:19-33

Corchon, L., and S. Wilkie. 1996. "Double Implementation of the Ratio Correspondence by a Market Mechanism." *Economic Design* 2:325-337.

Cornelli, F. 1996. "Optimal Selling Procedures with Fixed Costs." *Journal of Economic Theory* 71:1-30.

d'Aspremont, C., and L. Gérard-Varet. 1979. "Incentives and Incomplete Information." *Journal of Public Economics* 11:25-45.

Drèze and de la Vallée Poussin. 1971. "A Tâtonnement Process for Public Goods." *Review of Economic Studies* 38:133-50.

Else, P.K. 1981. "A Reformulation of the Theory of Optimal Congestion Taxes." *Journal of Transport Economics and Policy* 15:217-232.

Gardner, R., E. Ostrom, and J. Walker. 1990. "The Nature of Common-Pool Resource Problems." *Rationality and Society* 2:335-58.

Gardner, R., E. Ostrom, and J. Walker. 1992. "Covenants With and Without a Sword: Self-Governance is Possible." *American Political Science Review* 86:404-417.

Groves, T. 1973. "Incentives in Teams." *Econometrica* 41:617-663.

Groves, T., and J. Ledyard. 1977. "Optimal Allocation of Public Goods: A Solution to the Free Rider Problem." *Econometrica* 45:783-810.

Guesnerie, R., and J.-J. Laffont. 1984. "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm." *Journal of Public Economics* 25:329-69.

Holmström, B., and R. Myerson. 1983. "Efficient and Durable Decision Rules with Incomplete Information." *Econometrica* 51:1799-1819.

Hurwicz, L. 1979. "Outcome Functions Yielding Walrasian and Lindahl Allocations at Nash Equilibrium Points." *Review of Economic Studies* 46:217-25.

Ito, M., T. Saijo, and M. Une. 1995. "The Tragedy of the Commons Revisited Identifying Behavioral Principals." *Journal of Economic Behavior and Organization* 28:311-335.

Laffont, J.J., and E. Maskin. 1979. "A Differential Approach to Expected Utility Maximizing Mechanisms." In *Aggregation and Revelation of Preferences*, edited by J.J. Laffont, pp. 289-308. Amsterdam: North-Holland.

Laffont, J.J., and E. Maskin. 1982. "The Theory of Incentives: An Overview." In *Advances in Economic Theory*, edited by W. Hildenbrand, pp. 31-94. Cambridge: Cambridge University Press.

Ledyard, J., and T. Palfrey. 1999. "A Characterization of Interim Efficiency with Public Goods." *Econometrica* 67(2):435-448.

Ledyard, J., and T. Palfrey. 1996. "Voting is Approximately Optimal: A Characterization of Interim Efficiency with Public Goods." *Social Science Working Paper #956*, California Institute of Technology.

Ledyard, J., and T. Palfrey. 1994. "Voting and Lottery Drafts as Efficient Public Goods Mechanisms." *Review of Economic Studies* 61:327-355

Malinvaud. 1972. "A Planning Approach to the Public Good Problem." *Swedish Journal of Economics* 73:96-112.

Mirrlees, J.A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38:175-208.

Moulin, H. 1994. "Serial Cost-Sharing of Excludable Public Goods." *Review of Economic Studies* 61:305-325.

Moulin, H., and S. Shenker. 1992. "Serial Cost Sharing." *Econometrica* 60:1009-1037.

Myerson, R. 1981. "Optimal Auction Design." *Mathematics of Operations Research* 6:58-73.

Pauly, M. 1967. "Clubs, Commonality, and the Core: An Integration of Game Theory and the Theory of Public Goods." *Economica* 34:314-324.

Pauly, M. 1970. "Cores and Clubs." *Public Choice* 9:53-65.

Peleg, B. 1996. "Double Implementation of the Lindahl Equilibrium by a Continuous Mechanism." *Economic Design* 2:311-324.

Rochet, J.-C. 1987. "A Necessary and Sufficient Condition for Rationalizability in a Quasi-Linear Context." *Journal of Mathematical Economics* 16:191-200.

Scotchmer, S., and M.H. Wooders. 1987. "Competitive Equilibrium and the Core in Club Economies with Anonymous Crowding." *Journal of Public Economics* 34:159-173.

Sorenson, J.R., J.T. Tschirhart, and A.B. Whinston. 1978. "Private Good Clubs and the Core." *Journal of Public Economics* 10:77-95.

Tian, G. 1989. "Implementation of the Lindahl Correspondence by a Single-Valued, Feasible, and Continuous Mechanism." *Review of Economic Studies* 56:613-621.

Tideman, N. 1972. "The Efficient Provision of Public Goods." In *Public Prices for Public Products*, edited by S. Mushkin, Washington: The Urban Institute.

Verhoef, E.T., R.H.M. Emmerink, P. Nijkamp, and P. Rietveld. "Information Provision, Flat and Fine Congestion Tolling and the Efficiency of Road Usage." *Regional Science and Urban Economics* 26:505-529.

Vickrey, W. 1961. "Counterspeculation, Auctions, and Competitive Sealed Tenders." *Journal of Finance* 16:1-17.

Walker, M. 1981. "A Simple Incentive Compatible Scheme for Attaining Lindahl Allocations." *Econometrica* 49:65-71.

Wilson, R. 1993. "Design of Efficient Trading Procedures." In *The Double Auction Market Institutions: Theories and Evidence*, edited by D. Friedman and J. Rust. Proceedings of the Workshop on Double Auction Markets held June 1991 in Santa Fe, New Mexico. Proceeding Volume XIV. Addison-Wesley.

## Chapter III:

# In Defense of Unanimous Jury Verdicts: Mistrials, Communication, and Strategic Voting

## 1 INTRODUCTION

It is a widely held belief among legal theorists that the requirement of unanimous jury verdicts in criminal trials reduces the likelihood of convicting an innocent defendant. This belief is, to a large extent, dependent upon the assumption that all jurors will vote non-strategically -- that is, that jurors will not take strategic voting issues into consideration but that the jury decision will depend only upon interpretation of the evidence presented at trial. Recent literature, however, has suggested that the assumption of non-strategic voting by jurors may be inconsistent with Nash equilibrium behavior and has thus drawn into question the supposed benefits of unanimous jury verdicts.

The use of juries in criminal trials is based, at least in part, upon the belief that, when all individuals possess a common preference for selecting the "better" of two alternatives (in this case, conviction or acquittal), a group is more likely than any single individual to select the preferred option. This is the central argument behind the extensive literature that has developed based on Condorcet's Jury Theorem [Condorcet 1795/1976, Grofman and Feld 1988, Klevorick, Rothschild, and Winship 1984, Miller 1986, and Young 1988]. Analysis and extensions of this theorem have generally been statistical in nature, however, taking individual probabilities of correct decisions to be exogenously determined [Berg 1993; Ladha 1992, 1993, 1995]. An implicit element of this approach is the assumption that individuals behave in the same manner when they are acting as a dictator as when they are participating in a group decision process. In the framework of jury decision-making, this is equivalent to assuming that a juror's vote depends exclusively on her own private information (and perhaps shared public information) about the trial and does not depend upon considerations of strategic interaction within the jury.

In a recent paper, however, Austen-Smith and Banks [1996] illustrate that such non-strategic voting in group decisions may be inconsistent with Nash equilibrium behavior under fairly general conditions. In response, McLennan [1996] and Wit [1996] have attempted to rehabilitate the central notion of Condorcet's Jury Theorem, by identifying reasonable conditions under which Nash equilibrium behavior, though it may be inconsistent with non-strategic voting, still predicts that groups are more likely to make correct decisions than individuals.

Feddersen and Pesendorfer [1998] have adapted the general framework of Austen-Smith and Banks to the specific case of jury procedures in criminal trials and, in doing so, have derived some surprising results about unanimous jury verdicts. Feddersen and Pesendorfer construct a model of the jury process in which it is never a Nash equilibrium for all jurors to vote non-strategically under unanimity rule. Moreover, Nash equilibrium behavior in this model leads to higher probabilities of both convicting an innocent defendant and acquitting a guilty defendant under unanimity rule than under a wide variety of alternative voting rules, including simple majority rule. They conclude that, if their model is accurate, the societal objective of avoiding such jury errors may be better served by eliminating the requirement of unanimous verdicts in criminal cases.

The present paper extends the Feddersen and Pesendorfer model by adding certain minimal enhancements that we argue bring the model closer to actual jury procedures. In particular, we separately analyze the implications of (1) incorporating the possibility of mistrial and (2) allowing limited communication among jurors. Under each of these enhancements, we identify general conditions under which non-strategic voting is, in fact, a Nash equilibrium. We further demonstrate that under such voting equilibria, the conclusion of the inferiority of unanimous jury verdicts does not persist. That is, if the possibility of either mistrial or limited communication is introduced, it is no longer the case that unanimous jury verdicts generally produce equilibrium probabilities of convicting an innocent defendant and acquitting a guilty defendant that are higher than under alternative voting rules. Moreover, within the non-strategic voting equilibria that exist under these model enhancements, unanimity rule maximizes ex ante expected utility for all jurors.

## 2 THE BASIC MODEL

We first introduce the basic model of jury procedure which was analyzed by Feddersen and Pesendorfer and, more generally, by Austen-Smith and Banks. This model will serve as a point of departure and source of comparison for the new jury models introduced in this paper.

## 2.1 Basic Theoretical Framework

It is assumed that there are n jurors who will vote to determine the fate of a defendant. The set of jurors will be denoted by $N = \{1,2,...,n\}$ with an individual juror being represented by $j \in N$. There are two possible states of the world: the defendant is either guilty or innocent. We denote by G the state of the world in which the defendant is guilty and by I the state in which the defendant is innocent. The prior probability of state G is given by parameter r, with the prior probability of state I therefore being 1-r.

Note that Feddersen and Pesendorfer simplify the problem by assuming that the two states of the world occur with equal probability (r=0.5). While this assumption does not constitute a significant theoretical restriction, it does complicate the interpretation of the assumptions and results. In actual practice, it is likely that the value of r is greater than 0.5, considering the fact that criminal juries in federal courts, for example, find the defendant guilty in more than 80% of all cases [Vidmar, et al. 1997]. In the theoretical results of this paper, we will therefore allow the value of r to be variable, while also discussing the more simplified results that arise when r=0.5. In the specific examples presented, we will examine the cases r=0.5 and r=0.8.

In the basic model, there are two possible outcomes of the jury vote: the defendant is convicted, denoted C, or the defendant is acquitted, denoted A. Each juror can either vote for conviction (C) or acquittal (A). All votes are done by secret ballot and no abstentions are allowed. We will represent by $|C|$ the total number of votes for conviction and by $|A|$ the total number of votes for acquittal. In addition, $|C|_{-j}$ will denote the number of votes for conviction among all jurors other than j, or N/{j}, while $|A|_{-j}$ will denote the number of votes for acquittal among N/{j}.

A voting rule is described by a threshold $\hat{k}$, which is an integer between 0 and n. If $|C|$ $= \hat{k}$ the defendant is convicted, and the defendant is acquitted otherwise. Unanimity rule is represented by the voting rule $\hat{k} = n$, while simple majority rule is represented by the voting rule with $\hat{k}$ equal to the smallest integer greater than $^n/_2$.

The impact of the trial evidence is represented by a private signal received by each juror.. We will denote by $s_j$ the signal received by juror j. There are two possible signals, g or i, and the signal is correlated with the true state of the world. In particular, for all j, $\text{Prob}(s_j{=}g \mid G) = \text{Prob}(s_j{=}i \mid I) = p \in (0.5,1.0)$. Thus, the parameter p is the probability that a juror receives the "correct" signal (g in state G or i in state I) and 1-p is the probability that a juror receives the "incorrect" signal (i in state G or g in state I). We will denote by $|g|$ the total number of g signals received and by $|i|$ the total number of i signals received. In addition, $|g|_{-j}$ denotes the number of g signals among N/{j} while $|i|_{-j}$ denotes the number of i signals among N/{j}.

Note that, although juror signals are drawn independently given the true state of the world, they are correlated to each other in the sense that $\text{Prob}(s_j{=}g \mid s_i{=}g) = \text{Prob}(s_j{=}i \mid s_i{=}i)$ $= p^2+(1-p)^2 > 1/2 > 2p(1-p) = \text{Prob}(s_j{=}g \mid s_i{=}i) = \text{Prob}(s_j{=}i \mid s_i{=}g)$. In other words, juror i's signal provides her information about juror j's signal and, in particular, she believes that juror j is more likely to have a signal that matches her own signal than one that does not.

We will denote by $\beta(k,n)$ the posterior probability that the defendant is guilty conditional on k of n guilty signals:

$$\beta(k,n) = \frac{rp^k(1-p)^{n-k}}{rp^k(1-p)^{n-k} + (1-r)p^{n-k}(1-p)^k}$$

Let $u_j(O,S)$ be juror j's utility given outcome O in state S. It is assumed that $u_j(C,G) = u_j(A,I) = 0$, $u_j(C,I) = -q_j$, and $u_j(A,G) = -(1-q_j)$ where $q_j \in (0,1)$. Under this construction, any juror j will prefer conviction to acquittal whenever she believes the probability that the defendant is guilty is greater than $q_j$. In this sense, $1-q_j$ is a measure of what juror j considers to be "reasonable doubt."

Note that we should expect any juror j to have $q_j > 0.5$. To see this, recognize that the "more probable than not" standard of proof employed in most civil trials is equivalent to $q_j = 0.5$ for all $j \in N$. The "beyond a reasonable doubt" standard used in criminal trials, on the other hand, is a <u>strictly</u> higher standard of proof and therefore requires $q_j > 0.5$ for all $j \in N$. In particular, any juror j with $q_j < 0.5$ would prefer to convict even in some cases in which she believed the defendant was more likely innocent than guilty. While it is possible that such jurors exist, one of the specific purposes of the jury selection process is to eliminate candidates with such preferences. In the examples presented in this paper, we will therefore usually assume that $q_j \in (0.5, 1.0)$ for all $j \in N$.

Also note that the analysis of the basic model presented by Feddersen and Pesendorfer assumes common utilities for all jurors (i.e., $q_i = q_j$ for all $i,j \in N$), although this assumption may have been made purely for technical convenience. To assure the generality of the results of this paper, we will use individual utilities in all of the present analysis.

The behavior of a given juror j in the basic model is described by a strategy mapping, $\sigma_j$: $(0,1) \times \{g,i\} \rightarrow [0,1]$, with $\sigma_j(q_j, s_j)$ being the probability of voting to convict given utility parameter $q_j$ and signal $s_j$. Using this notation, we will define two different non-strategic voting strategies: *informative voting* and *sincere voting*.

Informative voting is defined as voting to convict whenever a guilty signal is received and voting to acquit whenever an innocent signal is received. In other words, to vote informatively is simply to "vote your signal" and thus honestly reveal your private information. The informative voting strategy for juror j is therefore given by:

$$\sigma_j(q_j, s_j) = \begin{cases} 1 \text{ if } s_j = g \\ 0 \text{ if } s_j = i \end{cases}$$

Note that informative voting is not only non-strategic, but also naive, since voting only according to one's signal may be inconsistent with expected utility maximization for some jurors. Thus, we also define sincere voting. A strategy for juror j is considered sincere voting when it consists of voting for the trial outcome which maximizes her expected utility conditional on her signal (and perhaps any other revealed signals). Thus, the general form of the sincere voting strategy for juror j is given by:

$$\sigma_j(q_j, s_j) = \begin{cases} 1 \text{ if } q_j < \Pr ob(G) \\ 0 \text{ if } q_j \geq \Pr ob(G) \end{cases}$$

We will contrast these non-strategic voting strategies with the strategic form of voting we call *rational voting*. Rational voting consists simply of voting according to Nash equilibrium behavior. Rational voting thus requires that a juror vote for the trial outcome which maximizes her expected utility conditional on her signal <u>and</u> conditional on her vote being pivotal; that is, that her vote can change the trial outcome. A rational juror must vote as if her vote is pivotal because this is the only case in which her vote will ever affect her utility. In the basic model, rational voting thus means voting as if exactly $\hat{k}-1$ other jurors are voting to convict.

Note that for a given voting rule there may be conditions under which rational voting is equivalent to informative and/or sincere voting; however, it may also be the case that these voting strategies do <u>not</u> coincide. One important result to recognize, however, is that whenever rational voting is equivalent to informative voting (i.e., whenever there exists a Nash equilibrium in which all jurors "vote their signal"), sincere voting will also be rational.

### 2.2 Assumptions and Conclusions of the Basic Model

In analyzing this basic model, Feddersen and Pesendorfer make several assumptions to eliminate potential equilibria that do not satisfy certain normative criteria. In particular, they eliminate from consideration asymmetric equilibria and equilibria in which a juror's strategy is independent of the signal received. Certain restrictions are also placed upon the relationship between the parameters p and $q_j$. In particular, it is assumed for all $j \in N$ that $1-p = q_j = \beta(n-1,n)$. The lower bound on $q_j$ here is not particularly restrictive, since it is generally assumed that $q_j$ is greater than 0.5 which is greater than 1-p. The upper bound on $q_j$ is also relatively permissive. This bound says only that n-1 guilty signals (versus only one innocent signal) is sufficient information for all jurors to prefer conviction.

With these assumptions placed on the basic model, Feddersen and Pesendorfer demonstrate that, under unanimity rule, there does not exist a Nash equilibrium in which all jurors vote informatively or sincerely and that there is instead a unique mixed strategy

Nash equilibrium. They also identify the unique Nash equilibrium for non-unanimous voting rules and illustrate that, as the size of the jury increases towards infinity, equilibrium behavior under unanimity rule leads to higher probabilities of both convicting an innocent defendant and acquitting a guilty defendant than under non-unanimous voting rules. Feddersen and Pesendorfer also use the following example to demonstrate that the inferiority of unanimous jury verdicts, while primarily a limit result, can also hold for smaller juries under fairly reasonable conditions.

**Example 1:** Let $n=12$, $r=0.5$, $p=0.8$, and $q_j=0.9$ for all $j \in N$. In this scenario, the probability of each type of trial error under different voting rules is given by the following chart:

| Voting Rule ($\hat{k}$) | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|
| Probability of Convicting the Innocent | 0.004 | 0.001 | 0.002 | 0.004 | 0.006 | 0.006 |
| Probability of Acquitting the Guilty | 0.019 | 0.066 | 0.135 | 0.245 | 0.420 | 0.654 |

Thus, the combined probability of either type of trial error is maximized under unanimity rule ($\hat{k}=12$) and minimized under simple majority rule ($\hat{k}=7$).

The key to understanding these somewhat surprising results is to recognize the significant influence that conditioning on being pivotal can have on juror strategies. For example, in the case of unanimity rule, conditioning on being pivotal means that each juror behaves as if <u>all</u> other jurors are voting to convict the defendant. It is therefore not difficult to see that, regardless of one's own signal, being pivotal provides a strong incentive to vote for conviction in this case since all other jurors are doing the same. For non-unanimous rules, on the other hand, being pivotal may provide much less compelling information: Under simple majority rule (with n odd), for example, being pivotal means only that an equal number of the other jurors are voting in each direction. This information is not overwhelming for either guilt or innocence, and can therefore be expected to have much less influence on juror voting.

To explicitly demonstrate that informative voting is not a Nash equilibrium under unanimity rule, suppose that all jurors do vote informatively and consider the situation in which juror j receives an innocent signal ($s_j$=i). It is easy to see that juror j has a positive incentive to deviate from informative voting and instead vote to convict in this case. First note that since juror j will condition her vote on being pivotal, she will behave as if all other jurors are voting to convict. When jurors vote informatively, this means that all other n-1 jurors received guilty signals and that juror j received the only innocent signal. Juror j's perceived probability of guilt is therefore $\beta$(n-1,n) in this case. However, by assumption, $q_j = \beta(n-1,n)$ and thus juror j prefers conviction to acquittal. Hence, juror j has an incentive to vote contrary to her own signal, and therefore informative voting is not a Nash equilibrium under unanimity rule in the basic model.

## 3 THE MISTRIAL MODEL

The first significant limitation of the basic model involves the delineation of trial outcomes. The basic model assumes that there are only two possible outcomes of the jury process: conviction or acquittal. Under unanimity rule, for example, a defendant is convicted if and only if all jurors vote for conviction, and the defendant is acquitted otherwise. In actual practice, however, almost all jurisdictions require unanimity to either convict or acquit a defendant in a criminal trial [Schwartz and Schwartz 1992]. If the jury vote results in neither a unanimous vote to convict nor a unanimous vote to acquit, then there is a "hung jury." If the hung jury situation persists through deliberations, a mistrial is declared and a new trial can be expected to take place. If the jury process is to be represented by a single vote, any non-unanimous vote would then immediately result in a mistrial.

### 3.1 Existence of Informative and Sincere Equilibria with Exogenous Mistrial Utilities

Thus, consider an enhancement to the basic model in which there are three possible outcomes of the jury process: the defendant is convicted (C), the defendant is acquitted (A), or a mistrial is declared (M). A voting rule is still described by an integer threshold $\hat{k}$. If $|C| = \hat{k}$ the defendant is convicted, if $|A| = \hat{k}$ the defendant is acquitted, and a mistrial is declared otherwise. Note that $\hat{k}$ must again be less than or equal to n but must now also be strictly greater than $\frac{n+1}{2}$. This lower bound on $\hat{k}$ exists in the mistrial model because if $\hat{k} = \frac{n+1}{2}$ then the trial outcome may be indeterminate in some cases or a mistrial may be an impossibility (which occurs when n is odd and $\hat{k} = \frac{n+1}{2}$).

Let $u_j(M,G) = -m_j^G$ $u_j(M,I) = -m_j^I$. We will make the natural assumption that the utility of a mistrial is strictly between the utilities of acquittal and conviction. That is, $0 < m_j^I < q_j$ and $0 < m_j^G < (1-q_j)$. In the next section, we will endogenize these mistrial utilities by equating them with the expected value of a new trial in a repeated trial process. For now, however, it is instructive to consider these mistrial utilities as exogenously determined.

Before proceeeding, it should be noted that Schwartz and Schwartz [1992] have also analyzed the impact of alternative voting rules within a model of jury procedure allowing for the possibility of mistrial. The Schwartz model, however, takes a very different approach, in which jurors have single-peaked preferences over a range of possible charges and the key choice variable is the prosecutorial decision about which charge (or charges) to prosecute.

The first result in the analysis of the current "mistrial model" presents the necessary and sufficient conditions for informative voting to be a Nash equilibrium.

**Proposition 1:** Informative voting is a Nash equilibrium in the mistrial model if and only if, for all jurors $j \in N$, the following two conditions are satisfied:

$$\big((1-q_j)rp - q_j(1-r)(1-p)\big)(1-p)^{2\hat{k}-n-1} + \big(m_j^G rp - m_j^I(1-r)(1-p)\big)\big(p^{2\hat{k}-n-1} - (1-p)^{2\hat{k}-n-1}\big) \geq 0$$
$$\big(q_j(1-r)p - (1-q_j)r(1-p)\big)(1-p)^{2\hat{k}-n-1} + \big(m_j^I(1-r)p - m_j^G r(1-p)\big)\big(p^{2\hat{k}-n-1} - (1-p)^{2\hat{k}-n-1}\big) \geq 0$$

**Proof:** See Appendix.

While the conditions of Proposition 1 may be difficult to interpret, the important point to recognize is that once the possibility of mistrial is introduced to the model, the conditions under which informative voting is a Nash equilibrium become much more permissive. In particular, informative voting Nash equilbria <u>do</u> exist under unanimity rule in the mistrial model while this was not true for the basic model. Moreover, the conditions of Proposition 1 are actually fairly general as will be discussed in more detail below.

Recognize that the conditions of Proposition 1 not only represent necessary and sufficient conditions for the existence of an informative voting Nash equilibrium, but they also represent sufficient conditions for the existence of a sincere voting Nash equilibrium. This is true because, as discussed previously, whenever informative voting constitutes a Nash equilibrum, sincere voting will also constitute a Nash equilibrium.

The key element in the proof of Proposition 1 that distinguishes the predictions of the mistrial model from the predictions of the basic model is the understanding of what it means to be pivotal in the two different models. To illustrate the distinction, consider the case of unanimity rule. In the basic model under unanimity rule, a juror is pivotal only when all other jurors are voting to convict. This provides a strong incentive to vote for conviction, even for those jurors who receive an innocent signal. In the mistrial model, on the other hand, a juror is pivotal in two different cases: when all other jurors are voting to convict and when all other jurors are voting to acquit. Moreover, given an innocent signal in the mistrial model, a juror will believe that it is more likely that all other jurors are voting to acquit than that all other jurors are voting to convict. This provides such a juror a greater incentive to vote informatively. The same is true for jurors who receive a guilty signal.

Although the conditions of Proposition 1 are fairly general, the structure of the inequalities in the proposition makes it difficult to immediately characterize all of the parameter values for which the proposition is satisfied. It is therefore helpful to examine more straightforward conditions that are simply sufficient (but not necessary) for informative voting and sincere voting to each be a Nash equilibrium in the mistrial model.

One set of such sufficient conditions is the following:

$$\frac{r}{1-r}\cdot\frac{1-p}{p} \leq \frac{q_j}{1-q_j} \leq \frac{r}{1-r}\cdot\frac{p}{1-p} \quad \text{and} \quad \frac{r}{1-r}\cdot\frac{1-p}{p} \leq \frac{m_j^I}{m_j^G} \leq \frac{r}{1-r}\cdot\frac{p}{1-p}.$$

These conditions indicate that informative voting and sincere voting each constitute a Nash equilibrium whenever: (a) the utility of the two "incorrect" trial outcomes (convicting the innocent and acquitting the guilty) are not significantly different, and (b) the utility of the two mistrial outcomes are not significantly different. Depending upon the value of p, these conditions can be very general or rather restrictive, but they nonetheless illustrate that many non-trivial parameter values will satisfy Proposition 1.

It is important to note, however, that there are many parameter values that satisfy the conditions of Proposition 1 yet do not satisfy the easy-to-understand sufficient conditions specified above. For example, consider the Feddersen and Pesendorfer example in which $n=12$, $r=0.5$, $p=0.8$, and $q_j=0.9$ for all j (note that this example violates the above conditions). Under unanimity rule in this case, the conditions of Proposition 1 reduce approximately to:

$$\frac{1}{4} < \frac{m_j^I}{m_j^G} < 4.$$

This means that informative or sincere voting will be a Nash equilibrium for this example so long as the utility (or disutility) of one mistrial outcome is not more than four times as large as the utility (or disutility) of the other mistrial outcome.

## 3.2 Existence of Informative and Sincere Equilibria with Endogenous Mistrial Utilities

To further develop this mistrial model, we would like to endogenize the mistrial utilities, $m_j^G$ and $m_j^I$, by specifying juror perceptions about the consequences of mistrial. These perceptions might incorporate many different factors, but it seems reasonable to model the utility of mistrial as simply the expected utility of an additional trial before a new jury.* In other words, we have:

$$\begin{aligned} m_j^G &= (1-q_j)\cdot \mathrm{Prob}_S(A \mid G) + m_j^G \cdot \mathrm{Prob}_S(M \mid G) \\ m_j^I &= q_j \cdot \mathrm{Prob}_S(C \mid I) + m_j^I \cdot \mathrm{Prob}_S(M \mid I) \end{aligned}$$

where $\mathrm{Prob}_S(O \mid S)$ is the probability of outcome O in a single trial when the true state is S.

When the utility of mistrial is specified in this manner, the conditions for the existence of a sincere voting Nash equilibrium are simplified significantly:

**Proposition 2:** Suppose the utility of mistrial is equal to the expected utility of an additional trial before a new jury. Informative voting is then a Nash equilibrium in the mistrial model for any voting rule $\hat{k}$ if and only if, for all $j \in N$, we have:

$$\frac{r}{1-r}\cdot\frac{1-p}{p} \le \frac{q_j}{1-q_j} \le \frac{r}{1-r}\cdot\frac{p}{1-p}$$

---

* We could also discount the expected utility of future trials or apply a fixed cost/disutility to each new trial. Mistrial utilities incorporating these factors still allow us to calculate refined necessary and sufficient conditions for sincere voting; however, analysis of such utility structures significantly increases the complexity of the presentation while providing minimal additional insight.

**Proof:** See Appendix.

As with Proposition 1, recognize that the inequality condition of Proposition 2 represents not only a necessary and sufficient condition for the existence of an informative voting Nash equilibrium, but also a sufficient condition for the existence of a sincere voting Nash equilibrium. Further recognize that Propositions 1 and 2 both suggest that the occurrence of informative and sincere voting among jurors may increase as the "accuracy" of trials improves. As p increases, and thus trials become more truth revealing, all of the conditions of Propositions 1 and 2 become easier to satisfy, and thus informative and sincere voting Nash equilibria will exist for more juries and more trials. If such non-strategic voting is a desirable outcome, this result provides an additional argument for legal reforms that may be expected to improve the likelihood that the true state of the world, guilt or innocence, is revealed at trial.

It is helpful to discuss further the inequality condition from Proposition 2, because this condition will appear again later in the paper. First note that the inequality in the proposition is equivalent to the following condition: $\beta(0,1) = q_j = \beta(1,1)$ for all $j \in N$. This constraint can be interpreted as the "one-man jury condition," because it is the same condition that would be required for a one-man jury (or, more appropriately, a presiding judge) to ever render a meaningful verdict. To see this, consider a jury consisting of a single juror j. If $q_j < \beta(0,1)$, then all defendants will be convicted, no matter which signal is received by juror j. Similarly, if $q_j > \beta(1,1)$, then all defendants will be acquitted, no matter which signal is received by juror j. Thus, for this one-man jury to ever render a meaningful verdict (i.e., one that varies depending upon what happens at trial), it must be the case that $\beta(0,1) = q_j = \beta(1,1)$.

The significance of Proposition 2 clearly depends upon its generality, and it is therefore natural to ask how likely it is that the conditions of the proposition will be satisfied. This is, of course, an empirical question and beyond the scope of the current paper; however, the following example should serve to demonstrate that the conditions of the proposition are met quite easily:

**Example 2:** Consider the jury selection process for a felony trial in the state of California. This process involves the selection of 12 jurors from a large set of candidates who are interviewed by both the prosecution and defense.

The defense has 10 peremptory challenges to dismiss candidates who they believe are the most likely to convict. In our model, this is equivalent to dismissing candidates with the lowest q values. Similarly, the prosecution has 10 peremptory challenges to dismiss candidates who they believe are the least likely to convict. In our model, this is equivalent to dismissing candidates with the highest q values.

There are also an unlimited number of dismissals for cause, which the judge uses to eliminate candidates whose probability of voting for conviction is deemed either unacceptably low or unacceptably high. Dismissals for cause would be used in our model, for example, to eliminate candidates whose q values were below 0.5 or too close to 1.0. Given this candidate dismissal process, we see that from the first 32 candidates <u>not</u> dismissed for cause, a jury of 12 members can be chosen.

Let r=0.8 and p =0.8. In this case, an informative voting Nash equilibrium exists if and only if $0.50 = q_j = 0.94$ for all $j \in N$. Suppose that the distribution of q values from which the candidates (not dismissed for cause) is drawn is uniform between 0.5 and 1.0. Thus the probability that any one candidate violates the inequality above is 0.12. This gives us:

$$
\begin{aligned}
\text{Prob}(\exists \text{ informative voting equilibrium}) &= \text{Prob}\left(\forall j \in N,\ 0.50 \le q_j \le 0.94\right) \\
&= \text{Prob}\left(\left|\left\{j \in N : q_j > 0.94\right\}\right| \le 10\right) \\
&= \sum_{k=0}^{10} b(k, 32, 0.12) = 99.9\%
\end{aligned}
$$

Thus, in this example, the conditions of Proposition 2 are almost always satisfied.

### 3.3 Comparison of Alternative Voting Rules

Once the existence of non-strategic voting Nash equilibria is established, it is important to compare the performance of alternative voting rules in terms equilibrium outcomes. One possible performance measure is the probability of a trial error, in other words, the probability of convicting an innocent defendant or acquitting a guilty defendant. Proposition 3 indicates that the probabilities of convicting an innocent defendant and acquitting a guilty defendant both decrease as $\hat{k}$, the number of votes required for a verdict, increases.

**Proposition 3:** Suppose that mistrial always results in a new trial and consider two voting rules, $\hat{k}_1$ and $\hat{k}_2$, with $\hat{k}_1 < \hat{k}_2$. If jurors vote informatively, then:

(1) The probability of convicting an innocent defendant is lower under voting rule $\hat{k}_2$ than under voting rule $\hat{k}_1$.

(2) The probability of acquitting a guilty defendant is lower under voting rule $\hat{k}_2$ than under voting rule $\hat{k}_1$.

**Proof:** See Appendix.

Note that Proposition 3 implies that the probability of trial error is uniquely minimized by unanimity rule and uniquely maximized by simple majority rule. This result is in stark contrast to the conclusions from analysis of the basic model, in which Nash equilibrium behavior produced higher probabilities of both convicting an innocent defendant and acquitting a guilty defendant under unanimity rule than under any non-unanimous voting rule.

Another reasonable measure of the performance of alternative voting rules is in terms of expected utility. Our final result for the mistrial model indicates that the expected utility for any juror increases as the number of votes required for a verdict increases.

**Proposition 4:** Suppose that the utility of mistrial is equal to the expected utility of an additional trial before a new jury and consider two voting rules, $\hat{k}_1$ and $\hat{k}_2$, with $\hat{k}_1 < \hat{k}_2$. If jurors vote informatively, then the ex ante expected utility for a juror is higher under voting rule $\hat{k}_2$ than under voting rule $\hat{k}_1$.

**Proof:** See Appendix.

This proposition indicates that unanimity rule again performs uniquely best among all voting rules, this time in terms of maximizing expected utility. Moreover, Proposition 4 also implies that simple majority rule is again the uniquely worst voting rule under this performance measure.

While both Proposition 3 and Proposition 4 specifically apply to the version of the mistrial model in which mistrial utilities are determined endogenously, it is important to note that the basic results (i.e., unanimity rule minimizing error and maximizing utility) also hold true when mistrial utilities are specified exogenously. However, the analysis in the exogenous utility case is rather simple, and the appropriate interpretation of the results is less clear.

## 4 THE COMMUNICATION MODEL

Recall that the basic model effectively rules out any communication among jurors in that the entire jury process is assumed to be a single vote in which each juror has no information about the beliefs of other jurors. In actual practice, on the other hand, the jury process involves a significant amount of communication and information revelation and there are often several "straw votes" taken during deliberations.

Let us therefore now consider a different enhancement to the basic model allowing for minimal communication among jurors. In particular, suppose that the jury takes a single non-binding straw vote before taking the final binding vote for conviction or acquittal. All jurors must vote to either convict (C) or acquit (A) in both the preliminary and final vote, and the number of preliminary votes cast for each outcome are announced prior to conducting the final vote. It is assumed that no communication other than casting the preliminary vote takes place.

Note that this enhancement to the model is not meant to represent actual deliberation procedures, but is nonetheless intended to show the significance of including communication in any model of the jury process. The incorporation of a single non-binding straw-vote will demonstrate that the addition of even the most minimal communication can significantly change the conclusions of the model analysis.

### 4.1 Existence of Informative and Sincere Equilibria

We start our analysis of this "communication model" by defining a non-strategic strategy profile appropriate for the distinctive voting framework of the model. The *sincere revelation strategy profile* for the communication model consists of each juror j voting according to the following guidelines:

(1) In the preliminary vote, juror j votes to convict iff $s_j$=g (informative voting);

(2) In the final vote, juror j votes to convict iff $\beta(k,n)= q_j$ where k is the number of votes to convict from the preliminary vote (sincere voting);

Our first result for the communication model identifies the necessary and sufficient conditions for the sincere revelation strategy profile to constitute a subgame perfect Nash equilibrium.

**Proposition 5:** Let the jurors be numbered such that $q_1 = q_2 = \ldots = q_{n-1} = q_n$. Then the sincere revelation strategy profile is a subgame perfect Nash equilibrium for a given voting rule $\hat{k}$ if and only if one of the following conditions is true:

(a) $0 = q_{\hat{k}} = \beta(0,n)$;

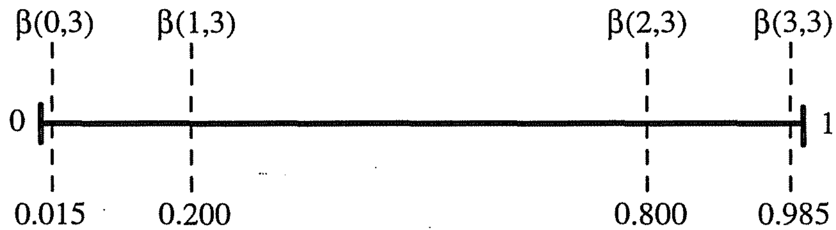
(b) $\beta(n,n) < q_{\hat{k}} = 1$; or

(c) $\exists k^* \in \{1, \ldots, n\}$ such that $\beta(k^*-1,n) = q_j = \beta(k^*,n)$ for all $j \in N$.

**Proof:** See Appendix.

Proposition 5 says that sincere revelation is a Nash equilibrium in the communication model whenever juror utilities satisfy a certain "closeness" condition. The basic insight behind this proposition is that when juror utilities are similar enough for there to be a situation of "common interest," everyone can benefit from an honest sharing of information in the preliminary vote. Since jurors do not have competing interests, the sharing of information can only serve to enhance the probability of achieving the outcome that all jurors prefer. In fact, the basic results of Proposition 5 should hold for any game of incomplete information and common interest in which a choice must be made between two alternatives, such as between two candidates for office or two public projects.

One simple situation that meets the conditions of the proposition is the case of common utilities (i.e., when $q_i$=$q_j$ for all $i,j \in N$). Thus, in Feddersen and Pesendorfer's example where $q_j$=0.9 for all j, sincere revelation voting is a Nash equilibrium in the communication model under all possible voting rules. It is important to note, however, that it is possible for juror utilities to differ significantly and still satisfy the conditions of the proposition. Consider the case of a three-person jury (n=3), and suppose a correct signal is received 80% of the time (p=0.8). In this case, we have $\beta(0,3)=0.015$, $\beta(1,3)=0.200$, $\beta(2,3)=0.800$, and $\beta(3,3)=0.985$, as shown in the figure below:




If all three $q_j$ values fall between any two of the dotted lines in this figure, Proposition 5 says that sincere revelation voting is a Nash equilibrium. It is thus clear that the $q_j$ values can differ significantly yet still satisfy the condition of the proposition.

It may seem that as n increases (i.e., the size of the jury becomes larger), the difference between $\beta(k^*-1,n)$ and $\beta(k^*,n)$ will become smaller for all $k^* \in \{1, ..., n\}$, making the conditions of Proposition 5 increasingly difficult to satisfy. This is not entirely true, however. In fact, some of these differences remain constant (and potentially rather large) for all values of n. Our next proposition uses this fact to identify sufficient conditions for the existence of a sincere voting equilibrium that are independent of the size of the jury.

**Proposition 6:** For n odd, the sincere revelation strategy profile is a subgame perfect Nash equilibrium for any voting rule $\hat{k}$ if, $\forall j \in N$:

$$\frac{r(1-p)}{r(1-p)+(1-r)p} \leq q_j \leq \frac{rp}{rp+(1-r)(1-p)}.$$

For n even, the sincere revelation strategy profile is a subgame perfect Nash equilibrium for any voting rule $\hat{k}$ if, $\forall j \in N$:

$$\frac{r(1-p)^2}{r(1-p)^2+(1-r)p^2} \leq q_j \leq r \quad \text{or} \quad r \leq q_j \leq \frac{rp^2}{rp^2+(1-r)(1-p)^2}.$$

**Proof:** See Appendix.

To better understand the scope of the conditions in Proposition 6, consider the case of r=0.5 with p=0.8. Proposition 2 then says that, if there is an odd number of jurors (whether there be 3 jurors, 11 jurors, or 99 jurors), the sincere revelation strategy profile will be a Nash equilibrium for any voting rule whenever $0.2 = q_j = 0.8$ for all jurors $j \in N$. In addition, if there is an even number of jurors (whether there be 4 jurors, 12 jurors, or 100 jurors) the sincere revelation strategy profile will be a Nash equilibrium for any voting rule whenever all juror utilities satisfy either $0.06 = q_j = 0.5$ or $0.5 = q_j = 0.94$. This example demonstrates that strategic jurors may vote sincerely in equilibrium under fairly general conditions for all juries and all voting rules.

To further illustrate the generality of Proposition 6, consider the following example:

**Example 3:** As in Example 2 above, consider again the jury selection process for a felony trial in the state of California. Recall that the defense has 10 peremptory challenges to dismiss candidates with the lowest q values and that the prosecution has 10 peremptory challenges to dismiss candidates with the highest q values. From the first 32 candidates (not dismissed for cause), a jury of 12 members can therefore be chosen.

Let r=0.5 and p =0.8. In this case, a sincere revelation Nash equilibrium exists whenever $0.50 \doteq q_j = 0.94$ for all $j \in N$. Again suppose that the distribution of q values from which the candidates is drawn is uniform between 0.5 and 1.0. Thus the probability that any one candidate violates the inequality above is 0.12. This gives us:

$$
\begin{aligned}
\text{Prob}(\exists \text{ sincere revelation equilibrium}) &> \Pr ob\left(\forall j \in N,\ 0.50 \leq q_j \leq 0.94\right) \\
&> \Pr ob\left(\left\{j \in N : q_j > 0.94\right\} \leq 10\right) \\
&> \sum_{k=0}^{10} b(k, 32, 0.12) = 99.9\%
\end{aligned}
$$

Thus, in this example, the conditions of Proposition 6 are almost always satisfied.

Note that Propositions 6 suggests that the occurrence of non-strategic voting among jurors in the communication model may increase as the “accuracy” of trials improves. We observed the same result in our analysis of the mistrial model. As p increases, and thus trials become more truth revealing, the conditions of Proposition 6 become easier to satisfy, and thus sincere voting Nash equilibria will exist for more juries and more trials. Also note that the condition in Proposition 6 for a jury with an odd number of members is equivalent to the “one-man jury condition” discussed previously.

Our next result for the communication model follows directly from Proposition 5.

**Proposition 7:** Suppose the juror utilities satisfy $0.5 = q_1 = q_2 = \ldots = q_{n-1} = q_n$. If condition (a), (b), or (c) from Proposition 5 is satisfied under voting rule $\hat{k}_1$, then the same condition is satisfied under any other voting rule $\hat{k}_2$ satisfying $\hat{k}_2 > \hat{k}_1$.

**Proof:** See Appendix.

This proposition indicates that, as long as $q_j$=0.5 for all j (as we would expect), sincere revelation voting is more likely to be a Nash equilibrium under unanimity rule than under any alternative voting rule.

### 4.2 Comparison of Alternative Voting Rules

We evaluate the performance of alternative voting rules in the communication model by once again examining the probability of trial error under different rules.

**Proposition 8:** Suppose that the sincere revelation strategy profile is a subgame perfect Nash equilibrium for two voting rules, $\hat{k}_1$ and $\hat{k}_2$. If jurors behave according to this Nash equilibrium, then:

(1) The probability of convicting an innocent defendant is the same under both voting rules.

(2) The probability of acquitting a guilty defendant is the same under both voting rules.

**Proof:** See Appendix.

Proposition 8 indicates that the sincere revelation Nash equilibrium results in the same probability of trial error under all voting rules. Thus, our conclusions once again contrast with the results from analysis of the basic model, in which unanimous jury verdicts were shown to be uniquely inferior under this performance measure.

Applying the alternative criterion of expected utility maximization, our results once again conflict with the negative assessment of unanimity rule from the analysis of the basic model. Instead, Proposition 9 indicates that the sincere revelation Nash equilibrium in the communication model produces the same expected utility under all voting rules.

**Proposition 9:** Suppose that the sincere revelation strategy profile is a subgame perfect Nash equilibrium for two voting rules, $\hat{k}_1$ and $\hat{k}_2$. If jurors behave according to this Nash equilibrium, then the expected utility for any juror is the same under both voting rules.

**Proof:** See Appendix.

## 5 CONCLUSIONS AND EXTENSIONS

Analysis of the basic model of jury procedure produces the somewhat surprising result that sincere voting can never be a Nash equilibrium under unanimity rule. Instead, a mixed strategy equilibrium exists in which unanimous jury verdicts are uniquely inferior in terms of minimizing the probability of trial error.

The objective of the current paper is to evaluate the impact that certain extensions of this basic model have on the existence of informative and sincere voting Nash equilibria. In particular, we examined the effects of introducing the possibility of mistrial and allowing limited communication upon the incentives for jurors to vote sincerely. In both cases, we find non-trivial conditions under which informative or sincere voting is indeed a Nash equilibrium. In addition, we compare the outcomes of these Nash equilibria under alternative voting rules and demonstrate that unanimity rule minimizes the probability of trial error and maximizes the ex ante expected utility of jurors.

An additional implication of the results of this paper is that the generality of sincere voting equilibria is strongly dependent upon the "accuracy" of trials. In particular, as the probability that the true state of the world is revealed at trial increases, the conditions for the existence of the informative or sincere voting Nash equilibria become more general in both the mistrial model and the communication model. This provides an additional argument in support of any legal reform that can be shown to produce more accurate impressions of guilt or innocence at trial.

While this paper was concerned only with the existence of pure strategy informative and sincere voting Nash equilibria, the investigation of the impacts of mistrial and communication should be extended to examine the existence and implications of mixed strategy and other non-sincere Nash equilibria. In particular, it is important to determine what happens when the conditions for existence of informative or sincere voting Nash equilibria that are identified in this paper are violated. Do the equilibria that exist in such situations still produce outcomes that make unanimity rule superior in terms of minimizing error and maximizing utility? Or do the results of the basic model prevail, with unanimity rule being outperformed by other voting rules such as simple majority rule?

A research approach that may be very helpful in addressing these questions as well as others would be to consider an information structure in which the q values for jurors are drawn from some known distribution function and each juror otherwise knows only her own q value. While the conditions for existence of non-strategic equilibria described in this paper encompass many of the parameter value combinations we might reasonably expect to observe, this alternative approach may produce results that are even more general.

An additional important extension of this research would be to identify the optimal jury institution by comparing alternatives that differ along several different dimensions, including the number of jurors, the voting rule employed, and the presence or absence of a mistrial outcome. In order to fully address this issue, however, one may need to specify a social welfare function that encompasses not only the utility of each possible trial outcome but also the social cost of multiple trials.

## APPENDIX

**Proposition 1:** Informative voting is a Nash equilibrium in the mistrial model if and only if, for all jurors $j \in N$, the following two conditions are satisfied:

$$\big(\big((1-q_j)rp - q_j(1-r)(1-p)\big)(1-p)^{2\hat{k}-n-1} + \big(m_j^G rp - m_j^I(1-r)(1-p)\big)\big(p^{2\hat{k}-n-1} - (1-p)^{2\hat{k}-n-1}\big)\big) \geq 0$$

$$\big(q_j(1-r)p - (1-q_j)r(1-p)\big)(1-p)^{2\hat{k}-n-1} + \big(m_j^I(1-r)p - m_j^G r(1-p)\big)\big(p^{2\hat{k}-n-1} - (1-p)^{2\hat{k}-n-1}\big) \geq 0$$

**Proof:** Recall that a strategic voter will condition her strategy on the event that her vote is pivotal. For a given juror j, there are exactly four scenarios in which her vote is pivotal:

(1) Defendant is guilty and $\hat{k}$-1 other jurors vote to convict $\left(G \cap |C|_{-j} = \hat{k}-1\right)$

(2) Defendant is guilty and $\hat{k}$-1 other jurors vote to acquit $\left(G \cap |A|_{-j} = \hat{k}-1\right)$

(3) Defendant is innocent and $\hat{k}$-1 other jurors vote to convict $\left(I \cap |C|_{-j} = \hat{k}-1\right)$

(4) Defendant is innocent and $\hat{k}$-1 other jurors vote to acquit $\left(I \cap |A|_{-j} = \hat{k}-1\right)$

Juror j's beliefs about the relative likelihood of each of these four scenarios will help determine her utility maximizing strategy. In particular, for any juror j, the expected utility of a vote to convict (ignoring the event in which the vote is not pivotal) is given by:

$$\begin{aligned}
EU_j(C, s_j) &= \Pr ob\left(G \cap |C|_{-j} = \hat{k}-1\right) u_j(C,G) + \Pr ob\left(G \cap |A|_{-j} = \hat{k}-1\right) u_j(M,G) \\
&\quad + \Pr ob\left(I \cap |C|_{-j} = \hat{k}-1\right) u_j(C,I) + \Pr ob\left(I \cap |A|_{-j} = \hat{k}-1\right) u_j(M,I) \\
&= -q_j \cdot \Pr ob\left(I \cap |C|_{-j} = \hat{k}-1\right) - m_j^G \cdot \Pr ob\left(G \cap |A|_{-j} = \hat{k}-1\right) \\
&\quad - m_j^I \cdot \Pr ob\left(I \cap |A|_{-j} = \hat{k}-1\right)
\end{aligned}$$

Similarly, the expected utility of a vote to acquit is given by:

$$\begin{aligned}EU_j(A,s_j) &= \Pr ob\left(G \cap |C|_{-j} = \hat{k}-1\right)\cdot u_j(M,G) + \Pr ob\left(G \cap |A|_{-j} = \hat{k}-1\right) u_j(A,G) \\ &\quad + \Pr ob\left(I \cap |C|_{-j} = \hat{k}-1\right) u_j(M,I) + \Pr ob\left(I \cap |A|_{-j} = \hat{k}-1\right) u_j(A,I) \\ &= -\left(1-q_j\right)\cdot \Pr ob\left(G \cap |A|_{-j} = \hat{k}-1\right) - m_j^G \cdot \Pr ob\left(G \cap |C|_{-j} = \hat{k}-1\right) \\ &\quad - m_j^I \cdot \Pr ob\left(I \cap |C|_{-j} = \hat{k}-1\right)\end{aligned}$$

Now suppose all jurors vote informatively. That is, $\sigma_j(q_j,g) = 1$ and $\sigma_j(q_j,i) = 0$ for all j, and thus $|C| = |g|$ and $|A| = |i|$. We must show that no juror can increase his or her utility by deviating from this strategy. More specifically, for all $j \in N$, we must show that:

(1) If $s_j = g$, then $EU_j(C,g) = EU_j(A,g)$

(2) If $s_j = i$, then $EU_j(A,i) = EU_j(C,i)$

<u>Case 1:</u> $s_j = g$

In this case, juror j's beliefs about the probability of the first scenario in which her vote is pivotal $\left(G \cap |C|_{-j} = \hat{k}-1\right)$ is given by:

$$\begin{aligned}\Pr ob\left(G \cap |C|_{-j} = \hat{k}-1 \,\middle|\, s_j = g\right) &= \text{Prob}\left(G \cap |g| = \hat{k} \,\middle|\, s_j = g\right) \\ &= \frac{\text{Prob}\left(G \cap |g| = \hat{k} \cap s_j = g\right)}{\text{Prob}\left(s_j = g\right)} \\ &= \frac{\text{Prob}(G)\cdot\text{Prob}\left(|g| = \hat{k} \,\middle|\, G\right)\cdot\text{Prob}\left(s_j = g \,\middle|\, |g| = \hat{k}\right)}{\text{Prob}(G)\cdot\text{Prob}\left(s_j = g \,\middle|\, G\right) + \text{Prob}(I)\cdot\text{Prob}\left(s_j = g \,\middle|\, I\right)}\end{aligned}$$

$$= \frac{r \cdot \frac{n!}{\hat{k}! \cdot \left(n-\hat{k}\right)} \cdot p^{\hat{k}}(1-p)^{n-\hat{k}} \cdot \frac{\hat{k}}{n}}{rp+(1-r)(1-p)}$$

$$= \frac{(n-1)!}{\left(\hat{k}-1\right) \cdot \left(n-\hat{k}\right)} \cdot \frac{r \cdot p^{\hat{k}}(1-p)^{n-\hat{k}}}{rp+(1-r)(1-p)}$$

$$= \Psi \cdot r \cdot p^{\hat{k}}(1-p)^{n-\hat{k}}$$

where $\Psi = \dfrac{(n-1)!}{\left(\hat{k}-1\right) \cdot \left(n-\hat{k}\right) \cdot (rp+(1-r)(1-p))}$

In the same manner, we can show that:

$$\Pr ob\left(G \cap |A|_{-j} = \hat{k}-1 \,\middle|\, s_j = g\right) = \Psi \cdot r \cdot p^{n-\hat{k}+1}(1-p)^{\hat{k}-1}$$
$$\Pr ob\left(I \cap |C|_{-j} = \hat{k}-1 \,\middle|\, s_j = g\right) = \Psi \cdot (1-r) \cdot p^{n-\hat{k}}(1-p)^{\hat{k}}$$
$$\Pr ob\left(I \cap |A|_{-j} = \hat{k}-1 \,\middle|\, s_j = g\right) = \Psi \cdot (1-r) \cdot p^{\hat{k}-1}(1-p)^{n-\hat{k}+1}$$

Thus, the expected utility of a vote to convict is given by:

$$\begin{aligned} EU_j(C,g) &= -q_j \Pr ob\left(I \cap |C|_{-j} = \hat{k}-1\right) - m_j^G \Pr ob\left(G \cap |A|_{-j} = \hat{k}-1\right) \\ &\quad - m_j^I \Pr ob\left(I \cap |A|_{-j} = \hat{k}-1\right) \\ &= -q_j \Psi (1-r) p^{n-\hat{k}}(1-p)^{\hat{k}} - m_j^G \Psi r p^{n-\hat{k}+1}(1-p)^{\hat{k}-1} - m_j^I \Psi (1-r) p^{\hat{k}-1}(1-p)^{n-\hat{k}+1} \\ &= -\Psi p^{n-\hat{k}}(1-p)^{n-\hat{k}}\left[q_j(1-r)(1-p)^{2\hat{k}-n} + m_j^G rp(1-p)^{2\hat{k}-n-1} + m_j^I (1-r) p^{2\hat{k}-n-1}(1-p)\right] \end{aligned}$$

Similarly, the expected utility of a vote to acquit is given by:

$$\begin{aligned} EU_j(A,g) &= -\left(1-q_j\right) \Pr ob\left(G \cap |A|_{-j} = \hat{k}-1\right) - m_j^G \Pr ob\left(G \cap |C|_{-j} = \hat{k}-1\right) \\ &\quad - m_j^I \Pr ob\left(I \cap |C|_{-j} = \hat{k}-1\right) \end{aligned}$$

$$= -(1-q_j)\Psi rp^{n-\hat{k}+1}(1-p)^{\hat{k}-1} - m_j^G\Psi rp^{\hat{k}}(1-p)^{n-\hat{k}} - m_j^I\Psi(1-r)p^{n-\hat{k}}(1-p)^{\hat{k}}$$
$$= -\Psi p^{n-\hat{k}}(1-p)^{n-\hat{k}}\left[(1-q_j)rp(1-p)^{2\hat{k}-n-1} + m_j^G rp^{2\hat{k}-n} + m_j^I(1-r)(1-p)^{2\hat{k}-n}\right]$$

We now show that condition (a) in Proposition 1 holds if and only if $EU_j(C,g) = EU_j(A,g)$:

$$\begin{aligned}
&\left((1-q_j)rp - q_j(1-r)(1-p)\right)(1-p)^{2\hat{k}-n-1} \\
&\quad + \left(m_j^G rp - m_j^I(1-r)(1-p)\right)\left(p^{2\hat{k}-n-1} - (1-p)^{2\hat{k}-n-1}\right) \geq 0 \\
\Leftrightarrow\ & (1-q_j)rp(1-p)^{2\hat{k}-n-1} + m_j^G rp\left(p^{2\hat{k}-n-1} - (1-p)^{2\hat{k}-n-1}\right) \\
&\quad \geq q_j(1-r)(1-p)^{2\hat{k}-n} + m_j^I(1-r)(1-p)\left(p^{2\hat{k}-n-1} - (1-p)^{2\hat{k}-n-1}\right) \\
\Leftrightarrow\ & (1-q_j)rp(1-p)^{2\hat{k}-n-1} + m_j^G rp^{2\hat{k}-n} + m_j^I(1-r)(1-p)^{2\hat{k}-n} \\
&\quad \geq q_j(1-r)(1-p)^{2\hat{k}-n} + m_j^G rp(1-p)^{2\hat{k}-n-1} + m_j^I p^{2\hat{k}-n-1}(1-p) \\
\Leftrightarrow\ & -\Psi p^{n-\hat{k}}(1-p)^{n-\hat{k}}\left[q_j(1-r)(1-p)^{2\hat{k}-n} + m_j^G rp(1-p)^{2\hat{k}-n-1} + m_j^I p^{2\hat{k}-n-1}(1-p)\right] \\
&\quad \geq -\Psi p^{n-\hat{k}}(1-p)^{n-\hat{k}}\left[(1-q_j)rp(1-p)^{2\hat{k}-n-1} + m_j^G rp^{2\hat{k}-n} + m_j^I(1-r)(1-p)^{2\hat{k}-n}\right] \\
\Leftrightarrow\ & EU_j(C,g) \geq EU_j(A,g)
\end{aligned}$$

<u>Case 2:</u> $s_j = i$

In this case, we can calculate juror j's beliefs about the relative probabilities of the four scenarios in which her vote is pivotal in the same manner as above. This gives us:

$$\begin{aligned}
\Pr ob\left(G \cap |C|_{-j} = \hat{k}-1 \middle| s_j = i\right) &= \Phi rp^{\hat{k}-1}(1-p)^{n-\hat{k}+1} \\
\Pr ob\left(G \cap |A|_{-j} = \hat{k}-1 \middle| s_j = i\right) &= \Phi rp^{n-\hat{k}}(1-p)^{\hat{k}} \\
\Pr ob\left(I \cap |C|_{-j} = \hat{k}-1 \middle| s_j = i\right) &= \Phi(1-r)p^{n-\hat{k}+1}(1-p)^{\hat{k}-1} \\
\Pr ob\left(I \cap |A|_{-j} = \hat{k}-1 \middle| s_j = i\right) &= \Phi(1-r)p^{\hat{k}}(1-p)^{n-\hat{k}}
\end{aligned}$$

where $\Phi = \dfrac{(n-1)!}{(\hat{k}-1)\cdot(n-\hat{k})\cdot(r(1-p)+(1-r)p)}$

Thus, the expected utility of a vote to convict is given by:

$$\begin{aligned}
EU_j(C,i) &= -q_j \Pr ob\left(I \cap |C|_{-j} = \hat{k}-1\right) - m_j^G \Pr ob\left(G \cap |A|_{-j} = \hat{k}-1\right) \\
&\quad - m_j^I \Pr ob\left(I \cap |A|_{-j} = \hat{k}-1\right) \\
&= -q_j\Phi(1-r)p^{n-\hat{k}+1}(1-p)^{\hat{k}-1} - m_j^G\Phi rp^{n-\hat{k}}(1-p)^{\hat{k}} - m_j^I\Phi(1-r)p^{\hat{k}}(1-p)^{n-\hat{k}} \\
&= -\Phi p^{n-\hat{k}}(1-p)^{n-\hat{k}}\left[q_j(1-r)p(1-p)^{2\hat{k}-n-1} + m_j^G r(1-p)^{2\hat{k}-n} + m_j^I(1-r)p^{2\hat{k}-n}\right]
\end{aligned}$$

Similarly, the expected utility of a vote to acquit is given by:

$$\begin{aligned}
EU_j(A,i) &= -\left(1-q_j\right) \Pr ob\left(G \cap |A|_{-j} = \hat{k}-1\right) - m_j^G \Pr ob\left(G \cap |C|_{-j} = \hat{k}-1\right) \\
&\quad - m_j^I \Pr ob\left(I \cap |C|_{-j} = \hat{k}-1\right) \\
&= -\left(1-q_j\right)\Phi rp^{n-\hat{k}}(1-p)^{\hat{k}} - m_j^G\Phi rp^{\hat{k}-1}(1-p)^{n-\hat{k}+1} - m_j^I\Phi(1-r)p^{n-\hat{k}+1}(1-p)^{\hat{k}-1} \\
&= -\Phi p^{n-\hat{k}}(1-p)^{n-\hat{k}}\left[\left(1-q_j\right)r(1-p)^{2\hat{k}-n} + m_j^G rp^{2\hat{k}-n-1}(1-p) + m_j^I(1-r)p(1-p)^{2\hat{k}-n-1}\right]
\end{aligned}$$

We now show that condition (b) in Proposition 1 holds if and only if $EU_j(A,i) = EU_j(C,i)$:

$$\begin{aligned}
&\left(q_j(1-r)p - \left(1-q_j\right)r(1-p)\right)(1-p)^{2\hat{k}-n-1} \\
&\quad + \left(m_j^I(1-r)p - m_j^G r(1-p)\right)\left(p^{2\hat{k}-n-1} - (1-p)^{2\hat{k}-n-1}\right) \geq 0 \\
&\Leftrightarrow\ q_j(1-r)p(1-p)^{2\hat{k}-n-1} + m_j^I(1-r)p\left(p^{2\hat{k}-n-1} - (1-p)^{2\hat{k}-n-1}\right) \\
&\quad \geq \left(1-q_j\right)r(1-p)^{2\hat{k}-n} + m_j^G r(1-p)\left(p^{2\hat{k}-n-1} - (1-p)^{2\hat{k}-n-1}\right)
\end{aligned}$$

$$\Leftrightarrow\ q_j(1-r)p(1-p)^{2\hat{k}-n-1} + m_j^G r(1-p)^{2\hat{k}-n} + m_j^I(1-r)p^{2\hat{k}-n}$$
$$\geq\ (1-q_j)r(1-p)^{2\hat{k}-n} + m_j^G r p^{2\hat{k}-n-1}(1-p) + m_j^I(1-r)p(1-p)^{2\hat{k}-n-1}$$
$$\Leftrightarrow\ -\Phi p^{n-\hat{k}}(1-p)^{n-\hat{k}}\left[(1-q_j)r(1-p)^{2\hat{k}-n} + m_j^G r p^{2\hat{k}-n-1}(1-p) + m_j^I(1-r)p(1-p)^{2\hat{k}-n-1}\right]$$
$$\geq\ -\Phi p^{n-\hat{k}}(1-p)^{n-\hat{k}}\left[q_j(1-r)p(1-p)^{2\hat{k}-n-1} + m_j^G r(1-p)^{2\hat{k}-n} + m_j^I(1-r)p^{2\hat{k}-n}\right]$$
$$\Leftrightarrow\ EU_j(A,i)\ \geq\ EU_j(C,i)$$

**Q.E.D.**

**Proposition 2:** Suppose the utility of mistrial is equal to the expected utility of an additional trial before a new jury. Informative voting is then a Nash equilibrium in the mistrial model for any voting rule $\hat{k}$ if and only if, for all $j \in N$, we have:

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

**Proof:** First note that, in a single trial, we have:

$$\Pr ob_S(C \mid I)\ =\ \Pr ob_S(A \mid G)\ =\ \sum_{x=\hat{k}}^{n}\binom{n}{x}p^{n-x}(1-p)^x$$
$$\Pr ob_S(M \mid I)\ =\ \Pr ob_S(M \mid G)\ =\ \sum_{x=n-\hat{k}+1}^{\hat{k}-1}\binom{n}{x}p^{n-x}(1-p)^x$$

This gives us:

$$m_j^G\ =\ (1-q_j)\cdot \text{Prob}_S(A \mid G) + m_j^G\cdot \text{Prob}_S(M \mid G)$$
$$m_j^G\ =\ (1-q_j)\sum_{x=\hat{k}}^{n}\binom{n}{x}p^{n-x}(1-p)^x + m_j^G\sum_{x=n-\hat{k}+1}^{\hat{k}-1}\binom{n}{x}p^{n-x}(1-p)^x$$

$$m_j^G = \frac{(1-q_j)\sum_{x=\hat{k}}^{n}\binom{n}{x}p^{n-x}(1-p)^x}{1-\sum_{x=n-\hat{k}+1}^{\hat{k}-1}\binom{n}{x}p^{n-x}(1-p)^x}$$

$$m_j^G = (1-q_j)\cdot\Omega$$

$$\text{where } \Omega = \frac{\sum_{x=\hat{k}}^{n}\binom{n}{x}p^{n-x}(1-p)^x}{1-\sum_{x=n-\hat{k}+1}^{\hat{k}-1}\binom{n}{x}p^{n-x}(1-p)^x} \geq 0$$

Similarly, we can show that:

$$m_j^I = q_j \cdot \Omega$$

Thus, condition (a) in Proposition 1 becomes:

$$\begin{aligned}
&\left((1-q_j)rp - q_j(1-r)(1-p)\right)(1-p)^{2\hat{k}-n-1} \\
&\qquad + \left(m_j^G rp - m_j^I(1-r)(1-p)\right)\left(p^{2\hat{k}-n-1} - (1-p)^{2\hat{k}-n-1}\right) \geq 0 \\
&\left((1-q_j)rp - q_j(1-r)(1-p)\right)(1-p)^{2\hat{k}-n-1} \\
&\qquad + \Omega\left((1-q_j)rp - q_j(1-r)(1-p)\right)\left(p^{2\hat{k}-n-1} - (1-p)^{2\hat{k}-n-1}\right) \geq 0 \\
&\left((1-q_j)rp - q_j(1-r)(1-p)\right)\left((1-p)^{2\hat{k}-n-1} + \Omega\left(p^{2\hat{k}-n-1} - (1-p)^{2\hat{k}-n-1}\right)\right) \geq 0 \\
&(1-q_j)rp - q_j(1-r)(1-p) \geq 0 \\
&\frac{q_j}{1-q_j} \leq \frac{r}{1-r}\cdot\frac{p}{1-p}
\end{aligned}$$

Similarly, we can show that condition (b) in Proposition 1 becomes:

$$\frac{q_j}{1-q_j} \geq \frac{r}{1-r}\cdot\frac{1-p}{p}$$

Thus, informative voting is a Nash equilibrium for any voting rule $\hat{k}$ if and only if, for all $j \in N$, we have:

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

**Q.E.D.**

**Proposition 3:** Suppose that mistrial always results in a new trial and consider two voting rules, $\hat{k}_1$ and $\hat{k}_2$, with $\hat{k}_1 < \hat{k}_2$. If jurors vote informatively, then:

(1) The probability of convicting an innocent defendant is lower under voting rule $\hat{k}_2$ than under voting rule $\hat{k}_1$.

(2) The probability of acquitting a guilty defendant is lower under voting rule $\hat{k}_2$ than under voting rule $\hat{k}_1$.

**Proof:** First note that, due to the symmetry of the mistrial model, the probability of convicting an innocent defendant is equal to the probability of acquitting a guilty defendant. Therefore, it is sufficient to prove only part (1) of the proposition.

In addition, note that it is sufficient to prove only that the probability of convicting an innocent defendant is lower under voting rule $\hat{k}_1+1$ than under voting rule $\hat{k}_1$. It is then obvious by induction that, for any voting rule $\hat{k}_2$ with $\hat{k}_1 < \hat{k}_2$, the probability of convicting an innocent defendant is lower under $\hat{k}_2$ than $\hat{k}_1$.

The probability of convicting an innocent defendant in (possibly) repeated trials under voting rule $\hat{k}_1$ is given by:

$$\mathrm{Prob}_R^{\hat{k}_1}(C \mid I) = \sum_{x=\hat{k}_1}^{n} \binom{n}{x} p^{n-x}(1-p)^x + \mathrm{Prob}(C \mid I)_R^{\hat{k}_1} \cdot \sum_{x=n-\hat{k}_1+1}^{\hat{k}_1-1} \binom{n}{x} p^{n-x}(1-p)^x$$

$$\mathrm{Prob}_R^{\hat{k}_1}(C \mid I)\left(1 - \sum_{x=n-\hat{k}_1+1}^{\hat{k}_1-1} \binom{n}{x} p^{n-x}(1-p)^x\right) = \sum_{x=\hat{k}_1}^{n} \binom{n}{x} p^{n-x}(1-p)^x$$

$$\mathrm{Prob}_R^{\hat{k}_1}(C \mid I) = \frac{\sum_{x=\hat{k}_1}^{n} \binom{n}{x} p^{n-x}(1-p)^x}{1 - \sum_{x=n-\hat{k}_1+1}^{\hat{k}_1-1} \binom{n}{x} p^{n-x}(1-p)^x}$$

$$\mathrm{Prob}_R^{\hat{k}_1}(C \mid I) = \frac{\sum_{x=\hat{k}_1}^{n} \binom{n}{x} p^{n-x}(1-p)^x}{\sum_{x=\hat{k}_1}^{n} \binom{n}{x} p^{n-x}(1-p)^x + \sum_{x=\hat{k}_1}^{n} \binom{n}{x} p^{x}(1-p)^{n-x}}$$

$$\mathrm{Prob}_R^{\hat{k}_1}(C \mid I) = \frac{\sum_{x=\hat{k}_1}^{n} \binom{n}{x} p^{n-x}(1-p)^x}{\sum_{x=\hat{k}_1}^{n} \binom{n}{x} \left(p^{n-x}(1-p)^x + p^x(1-p)^{n-x}\right)}$$

Similarly, the probability of convicting an innocent defendant in possibly repeated trials under voting rule $\hat{k}_1+1$ is given by:

$$\mathrm{Prob}_R^{\hat{k}_1+1}(C \mid I) = \frac{\sum_{x=\hat{k}_1+1}^{n} \binom{n}{x} p^{n-x}(1-p)^x}{\sum_{x=\hat{k}_1+1}^{n} \binom{n}{x} \left(p^{n-x}(1-p)^x + p^x(1-p)^{n-x}\right)}$$

We now show that $\mathrm{Prob}_R^{\hat{k}_1}(C \mid I) > \mathrm{Prob}_R^{\hat{k}_1+1}(C \mid I)$:

$$(1-p)^{2(x-\hat{k}_1)} < p^{2(x-\hat{k}_1)} \text{ for any } x > \hat{k}_1$$

$$p^{\hat{k}_1-x}(1-p)^{x-\hat{k}_1} < p^{x-\hat{k}_1}(1-p)^{\hat{k}_1-x} \text{ for any } x > \hat{k}_1$$

$$\binom{n}{x}p^{n+\hat{k}_1-x}(1-p)^{n+x-\hat{k}_1} < \binom{n}{x}p^{n+x-\hat{k}_1}(1-p)^{n+\hat{k}_1-x} \text{ for any } x > \hat{k}_1$$

$$\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}p^{n+\hat{k}_1-x}(1-p)^{n+x-\hat{k}_1} < \sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}p^{n+x-\hat{k}_1}(1-p)^{n+\hat{k}_1-x}$$

$$p^{\hat{k}_1}(1-p)^{n-\hat{k}_1}\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}p^{n-x}(1-p)^{x} < p^{n-\hat{k}_1}(1-p)^{\hat{k}_1}\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}p^{x}(1-p)^{n-x}$$

$$\left(p^{n-\hat{k}_1}(1-p)^{\hat{k}_1}+p^{\hat{k}_1}(1-p)^{n-\hat{k}_1}\right)\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}p^{n-x}(1-p)^{x}$$

$$< p^{n-\hat{k}_1}(1-p)^{\hat{k}_1}\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}\left(p^{n-x}(1-p)^{x}+p^{x}(1-p)^{n-x}\right)$$

$$\left(p^{n-\hat{k}_1}(1-p)^{\hat{k}_1}+p^{\hat{k}_1}(1-p)^{n-\hat{k}_1}+\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}\left(p^{n-x}(1-p)^{x}+p^{x}(1-p)^{n-x}\right)\right)\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}p^{n-x}(1-p)^{x}$$

$$< \left(p^{n-\hat{k}_1}(1-p)^{\hat{k}_1}+\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}p^{n-x}(1-p)^{x}\right)\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}\left(p^{n-x}(1-p)^{x}+p^{x}(1-p)^{n-x}\right)$$

$$\sum_{x=\hat{k}_1}^{n}\binom{n}{x}\left(p^{n-x}(1-p)^{x}+p^{x}(1-p)^{n-x}\right)\cdot\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}p^{n-x}(1-p)^{x}$$

$$< \sum_{x=\hat{k}_1}^{n}\binom{n}{x}p^{n-x}(1-p)^{x}\cdot\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}\left(p^{n-x}(1-p)^{x}+p^{x}(1-p)^{n-x}\right)$$

$$\frac{\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}p^{n-x}(1-p)^{x}}{\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}\left(p^{n-x}(1-p)^{x}+p^{x}(1-p)^{n-x}\right)} < \frac{\sum_{x=\hat{k}_1}^{n}\binom{n}{x}p^{n-x}(1-p)^{x}}{\sum_{x=\hat{k}_1}^{n}\binom{n}{x}\left(p^{n-x}(1-p)^{x}+p^{x}(1-p)^{n-x}\right)}$$

$$\mathrm{Prob}_R^{\hat{k}_1+1}(C \mid I) < \mathrm{Prob}_R^{\hat{k}_1}(C \mid I)$$

By induction, we have $\mathrm{Prob}_R^{\hat{k}_1}(C \mid I) > \mathrm{Prob}_R^{\hat{k}_2}(C \mid I)$, and since $\mathrm{Prob}_R(A \mid G) = \mathrm{Prob}_R(C \mid I)$, we also have that $\mathrm{Prob}_R^{\hat{k}_1}(A \mid G) > \mathrm{Prob}_R^{\hat{k}_2}(A \mid G)$.

**Q.E.D.**

**Proposition 4:** Suppose that the utility of mistrial is equal to the expected utility of an additional trial before a new jury and consider two voting rules, $\hat{k}_1$ and $\hat{k}_2$, with $\hat{k}_1 < \hat{k}_2$. If jurors vote informatively, then the ex ante expected utility for a juror is higher under voting rule $\hat{k}_2$ than under voting rule $\hat{k}_1$.

**Proof:** Note that it is sufficient to prove only that the ex ante expected utility is higher under voting rule $\hat{k}_1+1$ than under voting rule $\hat{k}_1$. It is then obvious by induction that, for any voting rule $\hat{k}_2$ with $\hat{k}_1 < \hat{k}_2$, the expected utility is higher under $\hat{k}_2$ than under $\hat{k}_1$.

If all jurors vote sincerely, the ex ante expected utility for juror j under voting rule $\hat{k}_1$ when the defendant is guilty is given by:

$$EU_j\left(\hat{k}_1 \middle| G\right) = -\left(1-q_j\right)\mathrm{Prob}_s^{\hat{k}_1}(A \mid G) + EU_j\left(\hat{k}_1 \middle| G\right)\mathrm{Prob}_s^{\hat{k}_1}(M \mid G)$$

$$EU_j\left(\hat{k}_1 \middle| G\right) = \frac{-\left(1-q_j\right)\mathrm{Prob}_s^{\hat{k}_1}(A \mid G)}{1-\mathrm{Prob}_s^{\hat{k}_1}(M \mid G)}$$

$$EU_j\left(\hat{k}_1 \middle| G\right) = \frac{-\left(1-q_j\right)\sum_{x=\hat{k}_1}^{n}\binom{n}{x}p^{n-x}(1-p)^x}{1-\sum_{x=n-\hat{k}_1+1}^{\hat{k}_1-1}\binom{n}{x}p^{n-x}(1-p)^x}$$

Similarly, we can show that the ex ante expected utility for juror j under voting rule $\hat{k}_1$ when the defendant is innocent is given by:

$$EU_j\left(\hat{k}_1 \middle| I\right) = \frac{-q_j\sum_{x=\hat{k}_1}^{n}\binom{n}{x}p^{n-x}(1-p)^x}{1-\sum_{x=n-\hat{k}_1+1}^{\hat{k}_1-1}\binom{n}{x}p^{n-x}(1-p)^x}$$

Thus, the overall ex ante expected utility for juror j under voting rule $\hat{k}_1$ is given by:

$$EU_j(\hat{k}_1) = \text{Prob}(G)\cdot\frac{-(1-q_j)\sum_{x=\hat{k}_1}^{n}\binom{n}{x}p^{n-x}(1-p)^x}{1-\sum_{x=n-\hat{k}_1+1}^{\hat{k}_1-1}\binom{n}{x}p^{n-x}(1-p)^x} + \text{Prob}(I)\cdot\frac{-q_j\sum_{x=\hat{k}_1}^{n}\binom{n}{x}p^{n-x}(1-p)^x}{1-\sum_{x=n-\hat{k}_1+1}^{\hat{k}_1-1}\binom{n}{x}p^{n-x}(1-p)^x}$$

$$= \frac{-(1-q_j)-q_j}{2}\cdot\frac{\sum_{x=\hat{k}_1}^{n}\binom{n}{x}p^{n-x}(1-p)^x}{1-\sum_{x=n-\hat{k}_1+1}^{\hat{k}_1-1}\binom{n}{x}p^{n-x}(1-p)^x}$$

$$= \frac{-\sum_{x=\hat{k}_1}^{n}\binom{n}{x}p^{n-x}(1-p)^x}{2-2\sum_{x=n-\hat{k}_1+1}^{\hat{k}_1-1}\binom{n}{x}p^{n-x}(1-p)^x}$$

Similarly, the overall ex ante expected utility for juror j under voting rule $\hat{k}_1+1$ is given by:

$$EU_j(\hat{k}_1+1) = \frac{-\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}p^{n-x}(1-p)^x}{2-2\sum_{x=n-\hat{k}_1}^{\hat{k}_1}\binom{n}{x}p^{n-x}(1-p)^x}$$

We now show that $EU_j(\hat{k}_1+1) > EU_j(\hat{k}_1)$:

$$(1-p)^{2(x-\hat{k}_1)} < p^{2(x-\hat{k}_1)} \quad \text{for any } x > \hat{k}_1$$

$$p^{\hat{k}_1-x}(1-p)^{x-\hat{k}_1} < p^{x-\hat{k}_1}(1-p)^{\hat{k}_1-x} \quad \text{for any } x > \hat{k}_1$$

$$\binom{n}{x}p^{n+\hat{k}_1-x}(1-p)^{n+x-\hat{k}_1} < \binom{n}{x}p^{n+x-\hat{k}_1}(1-p)^{n+\hat{k}_1-x} \quad \text{for any } x > \hat{k}_1$$

$$\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}p^{n+\hat{k}_1-x}(1-p)^{n+x-\hat{k}_1} < \sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}p^{n+x-\hat{k}_1}(1-p)^{n+\hat{k}_1-x}$$

$$\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}\left(p^{2n-\hat{k}_1-x}(1-p)^{\hat{k}_1+x}+p^{n+\hat{k}_1-x}(1-p)^{n+x-\hat{k}_1}\right)$$

$$<\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}\left(p^{2n-\hat{k}_1-x}(1-p)^{\hat{k}_1+x}+p^{n+x-\hat{k}_1}(1-p)^{n+\hat{k}_1-x}\right)$$

$$\binom{n}{\hat{k}_1}\left(p^{n-\hat{k}_1}(1-p)^{\hat{k}_1}+p^{\hat{k}_1}(1-p)^{n-\hat{k}_1}\right)\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}p^{n-x}(1-p)^{x}$$

$$<\binom{n}{\hat{k}_1}p^{n-\hat{k}_1}(1-p)^{\hat{k}_1}\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}\left(p^{n-x}(1-p)^{x}+p^{x}(1-p)^{n-x}\right)$$

$$\left(\sum_{x=\hat{k}_1}^{n}\binom{n}{x}\left(p^{n-x}(1-p)^{x}+p^{x}(1-p)^{n-x}\right)\right)\left(\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}p^{n-x}(1-p)^{x}\right)$$

$$<\left(\sum_{x=\hat{k}_1}^{n}\binom{n}{x}p^{n-x}(1-p)^{x}\right)\left(\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}\left(p^{n-x}(1-p)^{x}+p^{x}(1-p)^{n-x}\right)\right)$$

$$\frac{\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}p^{n-x}(1-p)^{x}}{\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}\left(p^{n-x}(1-p)^{x}+p^{x}(1-p)^{n-x}\right)}<\frac{\sum_{x=\hat{k}_1}^{n}\binom{n}{x}p^{n-x}(1-p)^{x}}{\sum_{x=\hat{k}_1}^{n}\binom{n}{x}\left(p^{n-x}(1-p)^{x}+p^{x}(1-p)^{n-x}\right)}$$

$$\frac{\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}p^{n-x}(1-p)^{x}}{1-\sum_{x=n-\hat{k}_1}^{\hat{k}_1}\binom{n}{x}p^{n-x}(1-p)^{x}}<\frac{\sum_{x=\hat{k}_1}^{n}\binom{n}{x}p^{n-x}(1-p)^{x}}{1-\sum_{x=n-\hat{k}_1+1}^{\hat{k}_1-1}\binom{n}{x}p^{n-x}(1-p)^{x}}$$

$$\frac{-\sum_{x=\hat{k}_1+1}^{n}\binom{n}{x}p^{n-x}(1-p)^{x}}{2-2\sum_{x=n-\hat{k}_1}^{\hat{k}_1}\binom{n}{x}p^{n-x}(1-p)^{x}}>\frac{-\sum_{x=\hat{k}_1}^{n}\binom{n}{x}p^{n-x}(1-p)^{x}}{2-2\sum_{x=n-\hat{k}_1+1}^{\hat{k}_1-1}\binom{n}{x}p^{n-x}(1-p)^{x}}$$

$$EU_j\left(\hat{k}_1+1\right)>EU_j\left(\hat{k}_1\right)$$

By induction, we have $EU_j\left(\hat{k}_2\right) > EU_j\left(\hat{k}_1\right)$.

**Q.E.D.**

**Proposition 5:** Let the jurors be numbered such that $q_1 = q_2 = ... = q_{n-1} = q_n$. Then the sincere revelation strategy profile is a subgame perfect Nash equilibrium for a given voting rule $\hat{k}$ if and only if one of the following conditions is true:

(a) $0 = q_{\hat{k}} = \beta(0,n)$;

(b) $\beta(n,n) < q_{\hat{k}} = 1$; or

(c) $\exists k^* \in \{1, ..., n\}$ such that $\beta(k^*-1,n) = q_j = \beta(k^*,n)$ for all $j \in N$.

**Proof:** First recognize that a strategic voter will condition her strategies in both the preliminary and final votes on the event that her vote is pivotal; that is, that her vote can change the trial outcome. In the event that her vote is <u>not</u> pivotal, her utility is unaffected by her vote and therefore such situations have no implications for strategic behavior.

We will evaluate strategy in the final vote first and then work backwards to examine the preliminary vote.

<u>Final Vote Strategy:</u>

Assume that in the preliminary vote, $\sigma_j(g)=1$ and $\sigma_j(i)=0$ for all jurors $j \in N$. Further assume that all jurors $j \in N$ vote to convict in the final vote iff $\beta(k,n) = q_j$, where k is the number of votes to convict from the preliminary vote. We must show that no juror has an incentive to deviate from this strategy in the final vote.

Note that, since all jurors vote sincerely in the preliminary vote, all jurors will know the total number of guilty (g) and innocent (i) signals before taking the final vote. Thus, all jurors will have the same estimate of the probability that the defendant is guilty, namely $\beta(k,n)$.

For any given juror, we need only consider the situation in which the juror's vote is pivotal. That is, if the given juror votes to convict, the defendant will be convicted, and if the given juror votes to acquit, the defendant will be acquitted. Thus, for any juror j, the expected utility of voting to acquit in this case is given by:

$$\begin{aligned} EU(A \mid |g| = k) &= -(1 - q_j) \cdot Prob(G \mid |g| = k) \\ &= -(1 - q_j) \cdot \beta(k,n) \end{aligned}$$

Similarly, the expected utility of voting to convict is given by:

$$\begin{aligned} EU(C \mid |g| = k) &= -q_j \cdot Prob(I \mid |g| = k) \\ &= -q_j \cdot (1 - \beta(k,n)) \end{aligned}$$

Therefore, juror j will want to vote to convict iff:

$$\begin{aligned} EU(C \mid |g| = k) &\geq EU(A \mid |g| = k) \\ -q_j \cdot Prob(I \mid |g| = k) &\geq -(1 - q_j) \cdot Prob(G \mid |g| = k) \\ -q_j \cdot (1 - \beta(k,n)) &\geq -(1 - q_j) \cdot \beta(k,n) \\ q_j \cdot (1 - \beta(k,n)) &\leq (1 - q_j) \cdot \beta(k,n) \\ q_j - q_j\beta(k,n) &\leq \beta(k,n) - q_j\beta(k,n) \\ q_j &\leq \beta(k,n) \end{aligned}$$

Therefore, sophisticated sincere voting in the final vote is a Nash equilibrium for this subgame. Recognize that this result is dependent only upon the assumption of sincere voting in the preliminary vote is independent of satisfaction or violation of conditions (a), (b), and (c).

Preliminary Vote Strategy:

Now assume that all jurors $j \in N$ vote to convict in the final vote iff $\beta(k,n) = q_j$, where k is the number of votes to convict from the preliminary vote. Further assume that $\sigma_j(g)=1$ and $\sigma_j(i)=0$ for all jurors $j \in N$ in the preliminary vote. We must show that, if one of the conditions, (a), (b), or (c), is satisfied, then no juror has an incentive to deviate from this sincere voting strategy in the preliminary vote. We must also show that, if all three conditions are violated, then at least one juror has an incentive to deviate from sincerity in the preliminary vote.

Case 1: Condition (a) is satisfied

In this case, we have that $0 = q_1 = \ldots = q_{\hat{k}} = \beta(0,n)$. This means that, in the final vote, at least $\hat{k}$ jurors will always vote to convict, and the defendant will thus always be convicted, regardless of the outcome of the preliminary vote. Therefore, no juror has a positive incentive to deviate from sincerity in the preliminary vote.

Case 2: Condition (b) is satisfied

In this case, we have that $\beta(n,n) = q_{\hat{k}} = \ldots = q_n = 1$ for all $j \in N$. This means that, in the final vote, at least $n-\hat{k}+1$ jurors will always vote to acquit, and the defendant will thus always be acquitted, regardless of the outcome of the preliminary vote. Therefore, no juror has a positive incentive to deviate from sincerity in the preliminary vote.

<u>Case 3:</u> Condition (c) is satisfied

In this case, we have that $\exists k^* \in \{1, ..., n\}$ such that $\beta(k^*-1,n) = q_j = \beta(k^*,n)$ for all $j \in N$. Thus, if juror j is pivotal in the preliminary vote, this means that $|g|_{-j}=k^*-1$. In other words, if juror j votes C in the preliminary vote, all other jurors will vote C in the final vote, and if juror j votes A in the preliminary vote, all other jurors will vote A in the final vote.

Note that this means that if juror j is pivotal in the preliminary vote, juror j can completely dictate the final trial outcome through her preliminary vote. Even under unanimity rule, juror j's preliminary vote will determine the final vote of all other jurors, thus allowing juror j to choose the trial outcome with her final vote. Thus, we can say that a juror will prefer to vote C in the preliminary vote if and only if she prefers that the defendant be convicted in the final outcome $\left(\text{i.e., } EU\left(C \mid |g|_{-j} = k^*-1\right) = EU\left(A \mid |g|_{-j} = k^*-1\right)\right)$.

Now suppose that $s_j=i$. In this case, we have that:

$$EU\left(C \mid |g|_{-j} = k^*-1\right) = EU\left(C \mid |g| = k^*-1\right) = -q_j \cdot (1-\beta(k^*-1,n))$$
$$EU\left(A \mid |g|_{-j} = k^*-1\right) = EU\left(A \mid |g| = k^*-1\right) = -(1-q_j)\cdot \beta(k^*-1,n)$$

$$\begin{aligned}\beta(k^*-1,n) < q_j &\Rightarrow \beta(k^*-1,n) - q_j \cdot \beta(k^*-1,n) < q_j - q_j \cdot \beta(k^*-1,n) \\ &\Rightarrow -(1-q_j)\cdot \beta(k^*-1,n) > -q_j \cdot (1-\beta(k^*-1,n)) \\ &\Rightarrow EU\left(A \mid |g|_{-j} = k^*-1\right) > EU\left(C \mid |g|_{-j} = k^*-1\right)\end{aligned}$$

Now suppose that $s_j$=g. In this case, we have that:

$$EU\left(C\,\middle|\,|g|_{-j}=k^*-1\right) = EU(C\,|\,|g|=k^*) = -q_j\cdot(1-\beta(k^*,n))$$
$$EU\left(A\,\middle|\,|g|_{-j}=k^*-1\right) = EU(A\,|\,|g|=k^*) = -(1-q_j)\cdot\beta(k^*,n)$$

$$\begin{aligned}
\beta(k^*,n) \geq q_j &\Rightarrow \beta(k^*,n)-q_j\cdot\beta(k^*,n) \geq q_j-q_j\cdot\beta(k^*,n)\\
&\Rightarrow (1-q_j)\cdot\beta(k^*,n) \geq q_j\cdot(1-\beta(k^*,n))\\
&\Rightarrow -(1-q_j)\cdot\beta(k^*,n) \leq -q_j\cdot(1-\beta(k^*,n))\\
&\Rightarrow EU\left(A\,\middle|\,|g|_{-j}=k^*-1\right) \leq EU\left(C\,\middle|\,|g|_{-j}=k^*-1\right)
\end{aligned}$$

Thus, a juror j will prefer to vote to convict in the preliminary vote if and only if $s_j$=g.

<u>Case 4:</u> Conditions (a), (b), and (c) are all violated

Violation of conditions (a) and (b) means that $\exists k^* \in \{1, ..., n\}$ such that $\beta(k^*-1,n) < q_{\hat{k}} = \beta(k^*,n)$. For a given juror j to be pivotal in the preliminary vote, it therefore means that $|g|_{-j}=k^*-1$. Violation of condition (c) means that $q_1 < \beta(k^*-1,n)$ and/or $\beta(k^*,n) < q_n$.

Suppose $q_1 < \beta(k^*-1,n)$ and consider the situation in which juror 1 is pivotal (i.e., $|g|_{-1}=k^*-1$) and $s_1$ = i. If juror 1 votes A in the preliminary vote (i.e., votes sincerely), the defendant will be acquitted, since $\beta(k^*-1,n) < q_{\hat{k}}$. However, if juror 1 instead deviates and votes C, the defendant will be convicted, since $q_{\hat{k}} = \beta(k^*,n)$. Since $q_1 < \beta(k^*-1,n)$, juror 1 prefers that the defendant is convicted, and therefore juror 1 has a positive incentive to deviate and vote C.

Now suppose $\beta(k^*,n) < q_n$ and consider the situation in which juror n is pivotal (i.e., $|g|_n = k^*-1$) and $s_n = g$. If juror n votes C in the preliminary vote (i.e., votes sincerely), the defendant will be convicted, since $q_{\hat{k}} = \beta(k^*,n)$. However, if juror n instead deviates and votes A, the defendant will be acquitted, since $\beta(k^*-1,n) < q_{\hat{k}}$. Since $\beta(k^*,n) < q_n$, juror n prefers that the defendant is acquitted, and therefore juror n has a positive incentive to deviate and vote A.

Thus, if conditions (a), (b), and (c) are all violated, then sincere voting is not a Nash equilibrium in the preliminary vote.

**Q.E.D.**

**Proposition 6:** For n odd, the sincere revelation strategy profile is a subgame perfect Nash equilibrium for any voting rule $\hat{k}$ if, $\forall j \in N$:

$$\frac{r(1-p)}{r(1-p)+(1-r)p} \leq q_j \leq \frac{rp}{rp+(1-r)(1-p)}.$$

For n even, the sincere revelation strategy profile is a subgame perfect Nash equilibrium for any voting rule $\hat{k}$ if, $\forall j \in N$:

$$\frac{r(1-p)^2}{r(1-p)^2+(1-r)p^2} \leq q_j \leq r \quad \text{or} \quad r \leq q_j \leq \frac{rp^2}{rp^2+(1-r)(1-p)^2}.$$

**Proof:** First, suppose that n is odd. Proposition 5 says that the sincere revelation strategy profile is a subgame perfect Nash equilibrium for any voting rule $\hat{k}$ if:

$$\beta(\tfrac{n-1}{2},n) \le q_j \le \beta(\tfrac{n+1}{2},n), \forall j \in N.$$

This condition is equivalent to:

$$\frac{rp^{\frac{n-1}{2}}(1-p)^{\frac{n+1}{2}}}{rp^{\frac{n-1}{2}}(1-p)^{\frac{n+1}{2}}+(1-r)p^{\frac{n+1}{2}}(1-p)^{\frac{n-1}{2}}} \le q_j \le \frac{rp^{\frac{n+1}{2}}(1-p)^{\frac{n-1}{2}}}{rp^{\frac{n+1}{2}}(1-p)^{\frac{n-1}{2}}+(1-r)p^{\frac{n-1}{2}}(1-p)^{\frac{n+1}{2}}}$$

$$\frac{r(1-p)}{r(1-p)+(1-r)p} \le q_j \le \frac{rp}{rp+(1-r)(1-p)}$$

Now, suppose that n is even. Proposition 5 says that the sincere revelation strategy profile is a subgame perfect Nash equilibrium for any voting rule $\hat{k}$ if:

$$\beta(\tfrac{n}{2}-1,n) \le q_j \le \beta(\tfrac{n}{2},n), \forall j \in N \quad \text{or} \quad \beta(\tfrac{n}{2},n) \le q_j \le \beta(\tfrac{n}{2}+1,n), \forall j \in N.$$

The first of these two conditions is equivalent to:

$$\frac{rp^{\frac{n}{2}-1}(1-p)^{\frac{n}{2}+1}}{rp^{\frac{n}{2}-1}(1-p)^{\frac{n}{2}+1}+(1-r)p^{\frac{n}{2}+1}(1-p)^{\frac{n}{2}-1}} \le q_j \le \frac{rp^{\frac{n}{2}}(1-p)^{\frac{n}{2}}}{rp^{\frac{n}{2}}(1-p)^{\frac{n}{2}}+(1-r)p^{\frac{n}{2}}(1-p)^{\frac{n}{2}}}$$

$$\frac{r(1-p)^2}{r(1-p)^2+(1-r)p^2} \le q_j \le r$$

The second of these two conditions is equivalent to:

$$\frac{rp^{\frac{n}{2}}(1-p)^{\frac{n}{2}}}{rp^{\frac{n}{2}}(1-p)^{\frac{n}{2}}+(1-r)p^{\frac{n}{2}}(1-p)^{\frac{n}{2}}} \le q_j \le \frac{rp^{\frac{n}{2}+1}(1-p)^{\frac{n}{2}-1}}{rp^{\frac{n}{2}+1}(1-p)^{\frac{n}{2}-1}+(1-r)p^{\frac{n}{2}-1}(1-p)^{\frac{n}{2}+1}}$$

$$r \le q_j \le \frac{rp^2}{rp^2+(1-r)(1-p)^2}$$

**Q.E.D.**

**Proposition 7:** Suppose the juror utilities satisfy $0.5 = q_1 = q_2 = \ldots = q_{n-1} = q_n$. If condition (a), (b), or (c) from Proposition 5 is satisfied under voting rule $\hat{k}_1$, then the same condition is satisfied under any other voting rule $\hat{k}_2$ satisfying $\hat{k}_2 > \hat{k}_1$.

**Proof:** Suppose condition (a) is satisfied for voting rule $\hat{k}_1$. This means that

$$0 = q_{\hat{k}_1} = \beta(0,n) = \frac{(1-p)^n}{p^n + (1-p)^n} < 0.5.$$

Since $q_j = 0.5$ for all j, condition (a) can not be satisfied for $\hat{k}_1$, and therefore the proposition is satisfied vacuously in this case.

Now suppose condition (b) is satisfied for voting rule $\hat{k}_1$. This means that $\beta(n,n) < q_{\hat{k}_1} = 1$. Since $\hat{k}_1 < \hat{k}_2$, we have that $q_{\hat{k}_1} < q_{\hat{k}_2}$, and thus that $\beta(n,n) < q_{\hat{k}_2} = 1$. Therefore, condition (b) is also satisfied for voting rule $\hat{k}_2$.

Finally, suppose that condition (c) is satisfied for voting rule $\hat{k}_1$. In this case, the condition is completely independent of the voting rule, thus condition (c) is also satisfied for voting rule $\hat{k}_2$.

**Q.E.D.**

**Proposition 8:** Suppose that the sincere revelation strategy profile is a subgame perfect Nash equilibrium for two voting rules, $\hat{k}_1$ and $\hat{k}_2$. If jurors behave according to this Nash equilibrium, then:

(1) The probability of convicting an innocent defendant is the same under both voting rules.

(2) The probability of acquitting a guilty defendant is the same under both voting rules.

**Proof:** Without loss of generality, assume $\hat{k}_1 = \hat{k}_2$. Existence of the sincere voting Nash equilibrium means that one of the three Proposition 5 conditions, (a), (b), or (c), is satisfied for each of the voting rules $\hat{k}_1$ and $\hat{k}_2$. It is also straightforward to show that both rules must satisfy the same condition (to see this, follow the same approach as used in the proof of Proposition 7).

Suppose both rules satisfy condition (a). In this case, $0 = q_j = \beta(0,n)$ for $j=1,2,\ldots,\hat{k}_1,\ldots,\hat{k}_2$. Thus, at least $\hat{k}_2$ jurors will always vote to convict in the final vote regardless of the outcome of the preliminary vote and regardless of the voting rule. Therefore, all defendants are convicted under both voting rules, and the probability of trial error under both voting rules is simply 0.5 (the prior probability that the defendant is innocent).

Now suppose both rules satisfy condition (b). In this case, $\beta(n,n) = q_j = 1$ for $j=\hat{k}_1,\ldots,\hat{k}_2,\ldots,n$. Thus, no more than $\hat{k}_1$ jurors will ever vote to convict in the final vote regardless of the outcome of the preliminary vote and regardless of the voting rule. Therefore, all defendants are acquitted under both voting rules, and the probability of trial error under both voting rules is simply 0.5 (the prior probability that the defendant is guilty).

Finally suppose both rules satisfy condition (c). In this case, $\exists k^* \in \{1, ..., n\}$ such that $\beta(k^*-1,n) = q_j = \beta(k^*,n)$ for all $j \in N$. Recall that the number of votes to convict in the preliminary vote will be equal to $|g|$ in equilibrium. Thus, if $|g|=k^*$, all jurors will vote to convict in the final vote, and if $|g|<k^*$, all jurors will vote to acquit in the final vote. Since all final votes are unanimous, if a defendant is convicted under one voting rule, she would also be convicted under the other voting rule. Therefore, the probability of convicting an innocent defendant must be the same under both voting rules. Similarly, if a defendant is acquitted under one voting rule, she would also be acquitted under the other voting rule. Therefore, the probability of acquitting a guilty defendant must also be the same under both voting rules.

**Q.E.D.**

**Proposition 9:** Suppose that the sincere revelation strategy profile is a subgame perfect Nash equilibrium for two voting rules, $\hat{k}_1$ and $\hat{k}_2$. If jurors behave according to this Nash equilibrium, then the expected utility for any juror is the same under both voting rules.

**Proof:** In the proof of Proposition 8, we showed that the trial outcome will always be the same under both voting rules. Therefore, the expected utility (and, in fact, the final realized utility) must be the same under both voting rules, also.

**Q.E.D.**

## REFERENCES

Austen-Smith, D., and J.S. Banks. 1996. "Information Aggregation, Rationality, and the Condorcet Jury Theorem." *American Political Science Review* 90(1):34-45.

Berg, S. 1993. "Condorcet's Jury Theorem, Dependency Among Jurors." *Social Choice and Welfare* 10:87-95.

Condorcet, M. de. 1785/1976. "Essay on the Application of Mathematics to the Theory of Decision-Making." Reprinted in *Condorcet: Selected Writings*, Keith Michael Baker, ed., 33-70. Indianapolis: Bobbs-Marrill Co.

Fedderson, T., and W. Pesendorfer. 1998. "Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting." *American Political Science Review* 92(1):23-36.

Grofman, B., and S.L. Feld. 1988. "Rousseau's General Will: A Condorcetian Perspective." *American Political Science Review* 82(2):567-576.

Klevorick, A.K., M. Rothschild, and C. Winship. 1984. "Information Processing and Jury Decisionmaking." *Journal of Public Economics* 23:245-278.

Ladha, K. 1992. "The Condorcet Jury Theorem, Free Speech, and Correlated Votes." *American Journal of Political Science* 36(3):617-634.

Ladha, K. 1993. "Condorcet's Jury Theorem in Light of De Finnetti's Theorem." *Social Choice and Welfare* 10:69-85.

Ladha, K. 1995. "Information Pooling Through Majority-Rule Voting: Condorcet's Jury Theorem with Correlated Votes." *Journal of Economic Behavior & Organization* 26:353-372.

Ladha, K., G. Miller, and J. Oppenheimer. 1996. "Information Aggregation by Majority Rule: Theory and Experiments." *Discussion Paper*, The John M. Olin School of Business, Washington University, St. Louis.

McLennan, A. 1996. "Consequences of the Condorcet Jury Theorem for Beneficial Information Aggregation by Rational Agents." *Discussion Paper,* Department of Economics, University of Minnesota.

Miller, N.R. 1986 "Information, Electorates, and Democracy: Some Extensions and Interpretations of the Condorcet Jury Theorem." *Information Pooling and Group Decision Making*, Bernard Grofman and Guillermo Owen, eds., 173-192. Greenwich: JAI Press Inc.

Schwartz, E.P., and W.F. Schwartz. 1992. "Decisionmaking by Juries Under Unanimity and Supermajority Voting Rules." *The Georgetown Law Journal* 80:775-807.

Vidmar, N., S.S. Beale, M. Rose, and L.F. Donnelly. 1997. "Should We Rush to Reform the Criminal Jury - Consider Conviction Rate Data." *Judicature* 80(6):286-290.

Wit, J. 1996. "Rational Choice and the Condorcet Jury Theorem." *Discussion Paper*, Division of Humanities and Social Sciences, California Institute of Technology.

Young, H.P. 1988. "Condorcet's Theory of Voting." *American Political Science Review* 82(4):1231-1244.

# Chapter IV:

# An Experimental Analysis of the Structure of Legal Fees: American Rule vs. English Rule

*With Charles R. Plott*

## 1 INTRODUCTION

The expanding volume of lawsuits and the ballooning of legal expenditures in recent years has attracted the interest, concern, and even anger of the American public and politicians. The number of lawsuits filed each year in the United States has grown steadily for several decades, with new filings in state and federal courts now approaching 19 million annually [The Economist 1992]. The American tort system is the most expensive in the world, with annual costs estimated at $117 billion [Hyde 1995]. Moreover, only about 40 cents from each dollar spent in this tort system actually serves to compensate victims while most of the rest pays for lawyer fees [O'Beirne 1995]. In addition, frequent examples of frivolous and outlandish suits in the popular media have also served to heighten public anger.

These developments have led law makers and legal professionals to consider alternative legal fee allocation rules as methods for administering justice more efficiently. Under the traditional American rule, parties to a lawsuit must each pay their own legal expenses.

One reform proposal is the English rule (also known as the British rule, “loser pays” rule, or indemnity system) under which the losing party must pay the prevailing party's attorney fees in addition to his or her own expenses. Both houses of Congress have recently passed legislation that mandates adoption of a form of the English rule in certain federal court cases.

Proponents of the English rule contend that its adoption would lead to fewer "frivolous" lawsuits and induce more of those suits that are filed to settle out of court. A change to the English rule, it is argued, would reduce the total volume of legal expenditure and eliminate the logjam of lawsuits that exists under the American rule. Nonetheless, there is considerable disagreement on whether or not application of the English rule would actually have these desired consequences in practice. As of yet, there is no consensus regarding the positive or negative effects of a change in legal fee allocation systems.

The implications of fee allocation rules are so widespread that any attempt to ascertain the full implications are far beyond the scope of this study. A narrowed focus is necessary. The four stages in the chronology of a legal dispute, as identified by Cooter and Rubinfeld [1989], are illustrated in Figure 1 (figures start on page IV-48) and will help provide a context for appropriately focusing the study. At every stage of a legal dispute, the parties involved make decisions that are influenced by their expectations of what might occur at subsequent stages of the dispute. As a result of this backward induction process, the entire system of behavior is heavily influenced by behavior at the (final) trial stage. Therefore, to fully understand the effects of different fee allocation rules on behavior and outcomes in legal disputes, a first investigation must focus on the effects at trial. Much of our research design reflects this objective.

The primary focus of this paper is on the different effects of the American rule and English rule on behavior and outcomes at trial. We study environments which can be interpreted as a legal procedure in which the probability of winning a lawsuit is partially determined by the relative legal expenditures of the plaintiff and defendant and partially determined by the inherent merits of the case. In addition to investigating trial decisions regarding legal expenditure, we also examine the effects of the two allocation rules on pretrial issues of suit and settlement.

The research poses four main questions. Do the two fee allocation rules have different effects on the level of legal expenditure? Do they have different effects on the frequency of suit, settlement, or trial? Are there any other factors that influence such differences in behavior? What are the best models for understanding the behavior and outcomes observed?

## 2 EXISTING RESEARCH

Previous research into the legal and social effects of different legal fee allocation rules has resulted in a wide variety of conclusions. These conclusions are often completely contradictory, particularly in the field of research regarding the effects on the frequency of suit, settlement, and trial. Several authors have concluded that a move from the American rule to the English rule would result in an increase in the number of suits being filed and an increase in the number of suits which proceed to trial [Shavell 1982, Bebchuk 1984, P'ng 1987, Donohue 1991b, Hylton 1993]. On the other hand, several others have concluded that such a move would instead decrease the number of suits and decrease the number of trials [Bowles 1987, Hause 1989, Hersch 1990, Spier 1994]. Still others have concluded that the number of suits and trials would necessarily be the same under both rules [Reinganum and Wilde 1986, Donohue 1991a] or that the effect of a change from one rule to the other would be ambiguous [Braeutigam, Owen, and Panzar 1984, Katz 1987, Gravelle 1993, Beckner and Katz 1995].

All models applied to understand the impact of alternative legal fee allocation rules are based on similar game-theoretic principles. However, the papers reach different conclusions, in part, because of the variety of conflicting (and sometimes restrictive) assumptions that are made by different researchers. The most significant assumption that has been made affecting this field of interest is that legal expenditures are fixed and exogenously determined. Under this assumption, litigants do not choose levels of legal expenditure and such expenditure does not influence trial outcome. Therefore, there are no strategic decisions or implications after a case has proceeded beyond settlement to trial. The fixed expenditure assumption is prevalent in the classic law and economics literature as well as recent analyses of fee allocation rules [Shavell 1982, Posner 1986, Reinganum and Wilde 1986, Coursey and Stanley 1988, Donohue 1991a, Gravelle 1993, Hylton 1993, Spier 1994].

Several authors have, however, incorporated the trial effects of legal expenditure into their examinations of fee allocation rules [Braeutigam, Owen, and Panzar 1984, Katz 1987, Plott 1987, Hause 1989, Hersch 1990], and these authors have universally concluded that legal expenditure at trial would be higher under the English rule than under the American rule. Nonetheless, these authors differ in their conclusions about the degree of difference in legal expenditure under the two rules, and agreement does not exist on the specific effects on plaintiff versus defendant expenditure.

An additional assumption that influences the conclusions in this field of research is that plaintiffs will bring suit if and only if they prefer trial to not filing suit [Shavell 1982, Hause 1989, Beckner and Katz 1995]. Such an assumption excludes consideration of forward looking plaintiffs who measure the expected gains from settlement as well as the expected gains from trial when considering whether or not to file suit. This assumption seems particularly troublesome when it is considered that at least 10 suits are settled out of court for every one suit that is resolved at trial [Boggs 1991].

The most significant empirical investigation of legal fee allocation rules has been conducted by Hughes and Snyder [1990, 1995], who examined trial data related to the State of Florida's temporary adoption of the English rule for medical malpractice legislation from 1980 to 1985. Hughes and Snyder concluded that the English rule produced significantly higher legal expenditure at trial but also reduced the number of trials by increasing the probability that claims would be dropped and increasing the likelihood of pretrial settlement for those claims that were not dropped. Plaintiff success rates at trial, average jury awards, and the value of out-of-court settlements were also all higher under the English rule than under the American rule.

Experimental research in the field of legal fee allocation mechanisms is very limited, although a few authors have done important work. Coursey and Stanley [1988] investigated the effect of legal fee allocation rules on pretrial bargaining, observing that the English rule tended to induce more settlements than the American rule. This work is limited, however, by the previously mentioned assumption of exogenously determined, fixed legal expenditures. Thomas [1994] incorporated the concept of endogenously chosen legal expenditures in an experimental investigation of the trial selection effect; however, this work is not directly related to the issue of legal fee allocation rules.

# 3 EXPERIMENTAL ENVIRONMENT AND PROCEDURE

This section introduces an experimental environment which can be interpreted as a legal dispute resolution procedure. This environment will facilitate an investigation of the different implications of the American and English rules.

## 3.1 NOTATION

The following notation will be necessary:

$A$ = amount of lawsuit

$C_P$ = fixed cost to plaintiff for bringing suit

$C_D$ = fixed cost to defendant for going to trial

$x_P$ = legal expenditure of plaintiff at trial

$x_D$ = legal expenditure of defendant at trial

$\alpha$ = relative productivity of lawyers in influencing legal outcome ( $0 \leq \alpha \leq 1$ )
(portion of outcome probability determined by legal expenditures)

$\pi$ = relative merit of plaintiff's case ( $0 \leq \pi \leq 1$ )
(probability plaintiff wins case in the absence of lawyer influence)

$P(x_P, x_D, \alpha, \pi)$ = probability that plaintiff wins the case

## 3.2 DEFINITION OF FEE ALLOCATION RULES

Applying the above notation, we can now formally define the American and English rules for allocation of legal fees.

*American Rule:* If the plaintiff wins the case at trial, the payoff to the plaintiff is $\Pi_P^A = A - C_P - x_P$ while the payoff to the defendant is $\Pi_D^A = -A - C_D - x_D$. If the defendant wins the case at trial, the payoff to the plaintiff is $\Pi_P^A = -C_P - x_P$ while the payoff to the defendant is $\Pi_D^A = -C_D - x_D$.

*English Rule:* If the plaintiff wins the case at trial, the payoff to the plaintiff is $\Pi_P^A = A - C_P$ while the payoff to the defendant is $\Pi_D^A = -A - C_D - x_P - x_D$. If the defendant wins the case at trial, the payoff to the plaintiff is $\Pi_P^A = -C_P - x_P - x_D$ while the payoff to the defendant is $\Pi_D^A = -C_D$.

## 3.3 LEGAL TECHNOLOGY

We will use a very explicit yet easily generalizable legal technology in this analysis. This technology is embodied in the function $P(x_P,x_D,\alpha,\pi)$, that is, the probability that the plaintiff wins the case. This probability is partially determined by the legal expenditures of the litigants (and therefore by the activity of lawyers) and partially determined by the inherent merits of the case. The specific functional form is as follows:

$$P(x_P, x_D, \alpha, \pi) = \alpha\left(\frac{x_P}{x_P + x_D}\right) + (1-\alpha)\pi$$

This function has several interesting properties:

- The probability the plaintiff prevails at trial is positively related to the merit of the case, $\pi$.

- For $\alpha > 0$, the probability the plaintiff wins increases as he increases his legal expenditure at trial. The same is true for the defendant.
- The marginal productivity of legal expenditure is given by $\dfrac{\partial P(x_P, x_D, \alpha, \pi)}{\partial x_P} = \dfrac{\alpha x_D}{(x_P + x_D)^2}$.
- The marginal productivity of legal expenditure increases as the productivity of lawyers, $\alpha$, increases.
- The marginal productivity of legal expenditure decreases as total legal expenditure, $x_P+x_D$, increases.
- Setting $\alpha = 0$ is equivalent to making the popular assumption that legal expenditure has no influence on trial outcome.
- For all values of $x_P$ and $x_D$, $(1-\alpha)\pi \leq P(x_P,x_D,\alpha,\pi) \leq \alpha+(1-\alpha)\pi$.

## 3.4 STRUCTURE OF THE EXPERIMENTAL LEGAL DISPUTE

The flow chart in Figure 2 illustrates the specific structure of the experimental legal dispute within which litigant behavior under the two alternative allocation rules is evaluated. During the actual experiments, neutral non-legal terminology is used to identify roles and actions; however, to avoid confusion, we use the equivalent legal terminology in the description that follows.

At the beginning of each legal dispute, every subject is randomly paired with another subject in the room. The identity of the persons they are paired with is never revealed to the subjects. After pairs are assigned, each member of each pair is randomly assigned a role, either plaintiff or defendant.

After roles are assigned, a level of $\pi$, or merit of the case, is randomly assigned to each pair. The three possible levels of $\pi$ are 0.25, 0.50, or 0.75. We will sometimes refer to a lawsuit with $\pi$=0.25 as a "frivolous" lawsuit, a lawsuit with $\pi$=0.50 as a "closely contested" lawsuit, and a lawsuit with $\pi$=0.75 as a "strong" lawsuit.

Next, each subject's role and merit is revealed to him or her. During the first series of experiments, Series 1, the assigned merit is revealed to the subjects with certainty. During Series 2, however, the merit is revealed with uncertainty, with each subject having a 60% chance of having the correct merit revealed to him or her and a 20% chance of having each of the other two incorrect merits revealed. For example, if a pair of subjects is assigned a merit of $\pi = 0.50$, each subject in the pair would have a 60% chance of being shown $\pi = 0.50$, a 20% chance of being shown $\pi = 0.25$, and a 20% chance of being shown $\pi = 0.75$.

Series 1 experiments will be referred to as "known merit" experiments while Series 2 experiments will be called "uncertain merit" experiments. The uncertain revelation of merit in the Series 2 experiments can be seen to represent incomplete discovery or imprecise communication between lawyer and client prior to trial. The subjects for the Series 2 experiments are selected from experienced subjects who have previously participated in Series 1 experiments.

After the revelation of roles and merits, the plaintiff in each pair is asked to choose whether to file suit or not file suit. If the plaintiff chooses to not file suit, the period ends for that pair and each receives a payoff of 0. If the plaintiff chooses to file suit, he incurs the fixed cost of $C_P$ for filing suit and the defendant is then asked whether she wants to settle or not settle.

In this experimental legal dispute, settlement means that the defendant simply pays the plaintiff the amount, A, for which the plaintiff is suing. We call this the "forfeiture settlement mechanism." This form of settlement is obviously extreme in the sense that no compromise is possible; however, this mechanism was chosen for several important reasons. First of all, theoretical and experimental analysis of the settlement bargaining process is a field of research without consensus about the proper model, and thus a somewhat arbitrary decision must be made when choosing a settlement mechanism. Moreover, in order to maintain adequate experimental control, we must employ a mechanism that minimizes the number of variables by limiting the interaction between litigants. The forfeiture settlement mechanism achieves this objective while still providing a reasonable opportunity for a significant number of disputes to be resolved prior to trial. Furthermore, although a restrictive mechanism may reduce the number of disputes settled, divergence in the frequency of settlement still provides valuable information about the different settlement incentives under the two alternative fee allocation rules. Lastly, since our primary interest is expenditure decisions at trial, we need to use a restrictive settlement mechanism to ensure that a sufficient number of legal disputes proceed to trial.

If the defendant chooses to settle, the plaintiff receives a payoff of $A-C_P$, while the defendant receives a payoff of $-A$. If the defendant chooses to not settle, the case proceeds to trial and each subject in the pair then chooses an amount, $x_P$ or $x_D$, to invest in legal expenditure at trial.

The probability that the plaintiff wins the case at trial is given by the legal technology function, $P(x_P,x_D,\alpha,\pi)$, specified above. The verdict is then determined by a random draw. If the plaintiff prevails at trial, he receives a payoff of $A-C_P-x_P$ under the American rule or $A-C_P$ under the English rule, while the defendant receives a payoff of $-A-C_D-x_D$ or $-A-C_D-x_P-x_D$ under the two rules respectively. If the defendant prevails at trial, she receives a payoff of $-C_D-x_D$ under the American rule or $-C_D$ under the English rule, while the plaintiff receives a payoff of $-C_P-x_P$ or $-C_P-x_P-x_D$ under the two rules respectively.

### 3.5 EXPERIMENTAL DESIGN PARAMETERS

A total of six experimental sessions were conducted with 10 or 12 students at the California Institute of Technology participating as subjects in each session. The experiments were conducted using a network of computers among the subjects, with subjects making decisions by pressing the appropriate keys on the keyboard.

The sessions are broken into 40 experimental periods, with each subject participating in a separate legal dispute each period. Half of all experimental disputes are conducted under the American rule, and half are conducted under the English rule.

During each experimental session, the productivity of lawyers, $\alpha$, is fixed at either 0.25 (low productivity), 0.50 (medium productivity), or 0.75 (high productivity). Two sessions have been conducted for each different level of lawyer productivity.

The currency used in the experiments is "francs," with five francs equivalent to one cent. Each experimental period, subjects receive a payment of 400 francs in addition to their payoff or loss from the legal dispute during the period. In all experimental sessions, the amount of the dispute, A, is set equal to 240 francs and the fixed costs, $C_P$ and $C_D$, are both set equal to 10 francs. In addition, the chosen levels of legal expenditure at trial, $x_P$

and $x_D$, are permitted to be any value between 0 and 1000 francs. In the end, the average cash payout for each experiment conducted was between 25 and 30 dollars per subject.

For additional clarification of the experimental environment and procedures, complete instructions and subject handouts from one experiment are included in the Appendix.

# 4 MODELS AND PREDICTIONS

In this section we discuss the predictions of behavior provided by the solution concepts of Nash equilibrium and subgame perfect equilibrium.

## 4.1 EXPECTED PROFIT FUNCTIONS

The definitions and legal technology function specified previously allow us to explicitly identify the expected profit function for each party when the legal dispute is to be resolved at trial. These expected profit functions will, of course, differ under the two alternative fee allocation rules.

Under the American rule, the expected profit for the plaintiff is given by:

$$\begin{aligned} E\Pi_P^A(x_P, x_D, \alpha, \pi) &= P(x_P, x_D, \alpha, \pi)A - x_P - C_P \\ &= A\alpha\left(\frac{x_P}{x_P + x_D}\right) + A(1-\alpha)\pi - x_P - C_P \end{aligned}$$

Similarly, the expected profit for the defendant under the American rule is given by:

$$E\Pi_D^A(x_P,x_D,\alpha,\pi) = P(x_P,x_D,\alpha,\pi)(-A) - x_D - C_D$$
$$= -A\alpha\left(\frac{x_P}{x_P+x_D}\right) - A(1-\alpha)\pi - x_D - C_D$$

Under the English rule, the expected profit for the plaintiff is given by:

$$E\Pi_P^E(x_P,x_D,\alpha,\pi) = P(x_P,x_D,\alpha,\pi)A + (1-P(x_P,x_D,\alpha,\pi))(-x_P-x_D) - C_P$$
$$= A - (A+x_P+x_D)(1-P(x_P,x_D,\alpha,\pi)) - C_P$$
$$= A - (A+x_P+x_D)\left(\alpha\left(\frac{x_D}{x_P+x_D}\right) + (1-\alpha)(1-\pi)\right) - C_P$$

Similarly, the expected profit for the defendant under the English rule is given by:

$$E\Pi_D^E(x_P,x_D,\alpha,\pi) = P(x_P,x_D,\alpha,\pi)(-A-x_P-x_D) - C_D$$
$$= -(A+x_P+x_D)\left(\alpha\left(\frac{x_P}{x_P+x_D}\right) + (1-\alpha)\pi\right) - C_D$$

## 4.2 MODEL PREDICTIONS: LEGAL EXPENDITURE AT TRIAL

**Proposition 1:** Under the American rule, if both parties are expected profit maximizers, the unique Nash equilibrium levels of legal expenditure at trial are:

$$x_P^A = x_D^A = \frac{A\alpha}{4}.$$

**Proof:** The plaintiff's objective is to

$$\max_{x_P}\left[A\alpha\left(\frac{x_P}{x_P+x_D}\right) + A(1-\alpha)\pi - x_P - C_P\right]$$

The first order condition is

$$A\alpha\left(\frac{x_D^A}{(x_P^A + x_D^A)^2}\right) - 1 = 0$$

$$A\alpha x_D^A = (x_P^A + x_D^A)^2$$

Similarly, solving the defendant's maximization problem, we get

$$A\alpha x_P^A = (x_P^A + x_D^A)^2$$

Combining these equations, we have

$$A\alpha x_P^A = A\alpha x_D^A$$

$$x_P^A = x_D^A$$

and thus

$$A\alpha x_P^A = (x_P^A + x_P^A)^2$$

$$A\alpha x_P^A = 4x_P^{A^2}$$

$$A\alpha = 4x_P^A$$

$$\frac{A\alpha}{4} = x_P^A = x_D^A$$

It is easily verified that these levels of expenditure at trial do indeed maximize the associated objective functions.

**Q.E.D.**

**Proposition 2:** Under the English rule, if both parties are expected profit maximizers, the unique Nash equilibrium levels of legal expenditure at trial are:

$$x_P^E = \frac{A\alpha\pi}{1-\alpha} \qquad \text{and} \qquad x_D^E = \frac{A\alpha(1-\pi)}{1-\alpha}.$$

**Proof:** The plaintiff's objective is to

$$\max_{x_P} \left[ A - (A + x_P + x_D)\left( \alpha\left( \frac{x_D}{x_P + x_D} \right) + (1-\alpha)(1-\pi) \right) - C_P \right]$$

The first order condition is

$$(A + x_P + x_D)\alpha\left( \frac{x_D}{(x_P + x_D)^2} \right) - \left( \alpha\left( \frac{x_D}{x_P + x_D} \right) + (1-\alpha)(1-\pi) \right) = 0$$

$$(A + x_P + x_D)\alpha x_D - \alpha x_D(x_P + x_D) - (1-\alpha)(1-\pi)(x_P + x_D)^2 = 0$$

$$A\alpha x_D - (1-\alpha)(1-\pi)(x_P + x_D)^2 = 0$$

$$\frac{A\alpha x_D}{(1-\alpha)(1-\pi)} = (x_P + x_D)^2$$

Similarly, solving the defendant's maximization problem, we get

$$\frac{A\alpha x_P}{(1-\alpha)\pi} = (x_P + x_D)^2$$

Combining these equations, we have

$$\frac{A\alpha x_D}{(1-\alpha)(1-\pi)} = \frac{A\alpha x_P}{(1-\alpha)\pi}$$

$$x_D = \frac{(1-\pi)x_P}{\pi}$$

Thus,

$$\frac{A\alpha x_P}{(1-\alpha)\pi} = \left(x_P + \frac{(1-\pi)x_P}{\pi}\right)^2$$

$$\frac{A\alpha x_P}{(1-\alpha)\pi} = \frac{x_P^{\,2}}{\pi^2}$$

$$x_P = \frac{A\alpha\pi}{1-\alpha}$$

and

$$x_D = \frac{(1-\pi)}{\pi}\left(\frac{A\alpha\pi}{1-\alpha}\right) = \frac{A\alpha(1-\pi)}{1-\alpha}$$

It is easily verified that these levels of expenditure at trial do indeed maximize the associated objective functions.

**Q.E.D.**

As further illustration of the Nash equilibrium predictions, Figure 3 illustrates the specific point predictions of legal expenditure at trial for the actual parameter values used in the experimental sessions.

## 4.3 MODEL PREDICTIONS: FORM OF RESOLUTION

To more clearly illustrate the predictions about the form of dispute resolution, we will assume in the following propositions that $C_P = C_D = C$. That is, we will assume that both parties face the same fixed costs, as is the case in the actual experimental sessions.

Allowing these fixed costs to differ does not qualitatively change the predictions; however, it adds unnecessary confusion.

We first note that in the trivial case in which C > A, the legal dispute will always be resolved with no lawsuit being filed. In other words, if the fixed costs of pursuing legal action exceed the possible gain for the plaintiff, she will never file suit. For this reason, the following propositions also assume that C is strictly less than A.

**Proposition 3:** Under the American rule, if both parties are expected profit maximizers, the unique subgame perfect equilibrium resolutions are as follows:

$$\text{(i) Settlement} \Leftrightarrow \pi > \frac{4-3\alpha-4\frac{C}{A}}{4(1-\alpha)}$$

$$\text{(ii) No Suit} \Leftrightarrow \pi \leq \min\left\{\frac{4-3\alpha-4\frac{C}{A}}{4(1-\alpha)},\ \frac{4\frac{C}{A}-\alpha}{4(1-\alpha)}\right\}$$

$$\text{(iii) Trial} \Leftrightarrow \text{Otherwise}$$

**Proof:** Combining the expected profit functions with the equilibrium trial expenditure predictions produces the following expected equilibrium profit functions under the American rule:

$$\begin{aligned} E\Pi_P^A(x_P,x_D,\alpha,\pi) &= A\alpha\left(\frac{x_P}{x_P+x_D}\right) + A(1-\alpha)\pi - x_P - C \\ &= A\alpha\left(\frac{1}{2}\right) + A(1-\alpha)\pi - \frac{A\alpha}{4} - C \\ &= A(\tfrac{1}{4}\alpha + \pi(1-\alpha)) - C \end{aligned}$$

$$
\begin{aligned}
E\Pi_D^A(x_P, x_D, \alpha, \pi) &= -A\alpha\left(\frac{x_P}{x_P + x_D}\right) - A(1-\alpha)\pi - x_D - C \\
&= -A\alpha\left(\frac{1}{2}\right) - A(1-\alpha)\pi - \frac{A\alpha}{4} - C \\
&= -A(\tfrac{3}{4}\alpha + \pi(1-\alpha)) - C
\end{aligned}
$$

Thus, the defendant strictly prefers settlement to trial if and only if

$$
\begin{aligned}
&-A(\tfrac{3}{4}\alpha + \pi(1-\alpha)) - C < -A \\
&\tfrac{3}{4}\alpha + \pi(1-\alpha) + \tfrac{C}{A} > 1 \\
&\pi(1-\alpha) > 1 - \tfrac{3}{4}\alpha - \tfrac{C}{A} \\
&4\pi(1-\alpha) > 4 - 3\alpha - 4\tfrac{C}{A} \\
&\pi > \frac{4 - 3\alpha - 4\frac{C}{A}}{4(1-\alpha)}
\end{aligned}
$$

Provided $C < A$, the plaintiff will always prefer settlement to no suit. Thus, whenever the above inequality holds, the unique subgame perfect equilibrium resolution is for the plaintiff to file suit and for the defendant to subsequently settle.

Note that a defendant who maximizes expected utility is actually indifferent between settlement and trial whenever $\pi = \dfrac{4-3\alpha-4\frac{C}{A}}{4(1-\alpha)}$. We have chosen to define the equilibrium choice of the defendant to be trial in this case, but note that we could have instead said that the defendant chooses settlement in this knife-edge situation. This would not change any of the substantive predictions of the model, and would simply require switching some strict inequalities to weak inequalities and vice versa (including changing the condition for no suit from a weak inequality to a strict inequality).

The plaintiff weakly prefers no suit to trial if and only if

$$A(\tfrac{1}{4}\alpha + \pi(1-\alpha)) - C \le 0$$
$$\tfrac{1}{4}\alpha + \pi(1-\alpha) - \tfrac{C}{A} \le 0$$
$$\pi(1-\alpha) \le \tfrac{C}{A} - \tfrac{1}{4}\alpha$$
$$4\pi(1-\alpha) \le 4\tfrac{C}{A} - \alpha$$
$$\pi \le \frac{4\frac{C}{A} - \alpha}{4(1-\alpha)}$$

Thus, the plaintiff prefers to not file suit whenever the defendant would not choose to settle and the above inequality holds. That is, the plaintiff will not file suit if and only if

$$\pi \le \min\left\{ \frac{4 - 3\alpha - 4\frac{C}{A}}{4(1-\alpha)},\ \frac{4\frac{C}{A} - \alpha}{4(1-\alpha)} \right\}$$

The legal dispute will obviously be resolved at trial whenever neither the conditions for settlement nor the conditions for no suit are met.

**Q.E.D.**

**Proposition 4:** Under the English rule, if both parties are expected profit maximizers, the unique subgame perfect equilibrium resolutions are as follows:

(i) Settlement $\Leftrightarrow \pi > (1-\tfrac{C}{A})(1-\alpha)$

(ii) No Suit $\Leftrightarrow \pi \le \min\left\{(1-\tfrac{C}{A})(1-\alpha),\ \tfrac{C}{A}(1-\alpha)+\alpha\right\}$

(iii) Trial $\Leftrightarrow$ Otherwise

**Proof:** Combining the expected profit functions with the equilibrium trial expenditure predictions produces the following expected equilibrium profit functions under the English rule:

$$\begin{aligned}
E\Pi_P^E(x_P, x_D, \alpha, \pi) &= A - (A + x_P + x_D)\left(\alpha\left(\frac{x_D}{x_P + x_D}\right) + (1-\alpha)(1-\pi)\right) - C \\
&= A - \left(A + \frac{A\alpha}{1-\alpha}\right)(\alpha(1-\pi) + (1-\alpha)(1-\pi)) - C \\
&= A - \left(\frac{A}{1-\alpha}\right)(1-\pi) - C \\
&= \frac{A - A\alpha - A + A\pi}{1-\alpha} - C \\
&= \frac{A(\pi - \alpha)}{1-\alpha} - C
\end{aligned}$$

$$\begin{aligned}
E\Pi_D^E(x_P, x_D, \alpha, \pi) &= -(A + x_P + x_D)\left(\alpha\left(\frac{x_P}{x_P + x_D}\right) + (1-\alpha)\pi\right) - C \\
&= -\left(A + \frac{A\alpha}{1-\alpha}\right)(\alpha\pi + (1-\alpha)\pi) - C \\
&= \frac{-A\pi}{1-\alpha} - C
\end{aligned}$$

Thus, the defendant strictly prefers settlement to trial if and only if

$$\begin{aligned}
&\frac{-A\pi}{1-\alpha} - C < -A \\
&\frac{\pi}{1-\alpha} + \tfrac{C}{A} > 1 \\
&\pi > (1 - \tfrac{C}{A})(1-\alpha)
\end{aligned}$$

Provided C < A, the plaintiff will always prefer settlement to no suit. Thus, whenever the above inequality holds, the unique subgame perfect equilibrium resolution is for the plaintiff to file suit and for the defendant to subsequently settle.

Note that a defendant who maximizes expected utility is actually indifferent between settlement and trial whenever $\pi = (1-\frac{C}{A})(1-\alpha)$. We have chosen to define the equilibrium choice of the defendant to be trial in this case, but note that we could have instead said that the defendant chooses settlement in this knife-edge situation. This would not change any of the substantive predictions of the model, and would simply require switching some strict inequalities to weak inequalities and vice versa (including changing the condition for no suit from a weak inequality to a strict inequality).

The plaintiff weakly prefers no suit to trial if and only if

$$\frac{A(\pi-\alpha)}{1-\alpha} - C \leq 0$$
$$\frac{\pi-\alpha}{1-\alpha} - \frac{C}{A} \leq 0$$
$$\pi \leq \frac{C}{A}(1-\alpha)+\alpha$$

Thus, the plaintiff prefers to not file suit whenever the defendant would not choose to settle and the above inequality holds. That is, the plaintiff will not file suit if and only if

$$\pi \leq \min\left\{\left(1-\frac{C}{A}\right)(1-\alpha),\ \frac{C}{A}(1-\alpha)+\alpha\right\}$$

The legal dispute will obviously be resolved at trial whenever neither the conditions for settlement nor the conditions for no suit are met.

**Q.E.D.**

As further illustration of the subgame perfect equilibrium predictions, Figures 4 and 5 illustrate the form of resolution predictions for the actual parameter values used in the experimental sessions.

## 4.5 OBSERVATIONS ABOUT MODEL PREDICTIONS

**Observation 1.** For $\alpha \in [0,1)$, total equilibrium trial expenditures under the English rule are <u>always</u> at least twice as large as the total equilibrium trial expenditures under the American rule:

$$
\begin{aligned}
x^A_{Total} &= x^A_P + x^A_D = \frac{A\alpha}{4} + \frac{A\alpha}{4} = \frac{A\alpha}{2} \\
x^E_{Total} &= x^E_P + x^E_D = \frac{A\alpha\pi}{(1-\alpha)} + \frac{A\alpha(1-\pi)}{(1-\alpha)} = \frac{A\alpha}{(1-\alpha)} \\
0 \le \alpha \le 1 &\Rightarrow 1-\alpha \le 1 \Rightarrow \frac{1}{1-\alpha} \ge 1 \Rightarrow \frac{A\alpha}{1-\alpha} \ge A\alpha
\end{aligned}
$$

$$\text{Thus, } x^E_{Total} \ge 2 \cdot x^A_{Total}$$

**Observation 2.** For $\alpha \in (0,1)$, equilibrium trial expenditure for the plaintiff is higher under the English rule than under the American rule iff $\pi > \frac{1}{4} - \frac{\alpha}{4}$, while equilibrium trial expenditure for the defendant is higher under the English rule than under the American rule iff $\pi < \frac{3}{4} + \frac{\alpha}{4}$:

$$\pi > \frac{1}{4} - \frac{\alpha}{4} \Leftrightarrow \frac{\pi}{1-\alpha} > \frac{1}{4} \Leftrightarrow \frac{A\alpha\pi}{1-\alpha} > \frac{A\alpha}{4} \Leftrightarrow x^E_P > x^A_P$$

$$\pi < \frac{3}{4} + \frac{\alpha}{4} \Leftrightarrow 1-\pi > \frac{1}{4} - \frac{\alpha}{4} \Leftrightarrow \frac{1-\pi}{1-\alpha} > \frac{1}{4} \Leftrightarrow \frac{A\alpha(1-\pi)}{1-\alpha} > \frac{A\alpha}{4} \Leftrightarrow x^E_D > x^A_D$$

Note that the above observation implies that, whenever $0.25 < \pi < 0.50$, equilibrium trial expenditures for both the plaintiff and defendant are higher under the English rule than under the American rule for any $\alpha \in (0,1)$.

**Observation 3.** Under both fee allocation rules with $\pi \in (0,1)$, equilibrium legal expenditure at trial increases as the productivity of lawyers increases:

$$\frac{\partial x_P^A}{\partial \alpha} = \frac{\partial x_D^A}{\partial \alpha} = \frac{A}{4} > 0$$

$$\frac{\partial x_P^E}{\partial \alpha} = \frac{A\pi}{(1-\alpha)^2} > 0$$

$$\frac{\partial x_D^E}{\partial \alpha} = \frac{A(1-\pi)}{(1-\alpha)^2} > 0$$

**Observation 4.** Under the American rule, equilibrium legal expenditure at trial is always no greater than one-fourth the amount of the suit:

$$x_P^A = x_D^A = \frac{A\alpha}{4} \leq \frac{A}{4} \qquad \forall \alpha \in [0,1]$$

**Observation 5.** Under the English rule with $\pi \in (0,1)$, equilibrium legal expenditure at trial increases without bound as the productivity of lawyers increases:

$$\lim_{\alpha \to 1} x_P^E = \lim_{\alpha \to 1} \frac{A\alpha\pi}{1-\alpha} = \infty \qquad \forall \pi \in (0,1]$$

$$\lim_{\alpha \to 1} x_D^E = \lim_{\alpha \to 1} \frac{A\alpha(1-\pi)}{1-\alpha} = \infty \qquad \forall \pi \in (0,1]$$

**Observation 6.** Under the American rule, equilibrium trial expenditure is independent of the merit of the case:

$$\frac{\partial x_P^A}{\partial \pi} = \frac{\partial x_D^A}{\partial \pi} = \frac{\partial}{\partial \pi}\left(\frac{A\alpha}{4}\right) = 0$$

**Observation 7.** Under the English rule with $\alpha \in (0,1)$, as the merit of the case increases, the equilibrium trial expenditure of the plaintiff increases and the equilibrium trial expenditure of the defendant decreases:

$$\frac{\partial x_P^E}{\partial \pi} = \frac{A\alpha}{1-\alpha} > 0$$
$$\frac{\partial x_D^E}{\partial \pi} = \frac{-A\alpha}{1-\alpha} < 0$$

**Observation 8.** In equilibrium under the English rule with $\alpha \in (0,1)$, (a) plaintiff expenditure at trial is less than defendant expenditure at trial iff $\pi<0.50$, (b) plaintiff expenditure at trial is equal to defendant expenditure at trial iff $\pi=0.50$, and (c) plaintiff expenditure at trial is greater than defendant expenditure at trial iff $\pi>0.50$:

$$x_P^E < x_D^E \Leftrightarrow \frac{A\alpha\pi}{1-\alpha} < \frac{A\alpha(1-\pi)}{1-\alpha} \Leftrightarrow \pi < 1-\pi \Leftrightarrow 2\pi < 1 \Leftrightarrow \pi < 0.50$$
$$x_P^E = x_D^E \Leftrightarrow \frac{A\alpha\pi}{1-\alpha} = \frac{A\alpha(1-\pi)}{1-\alpha} \Leftrightarrow \pi = 1-\pi \Leftrightarrow 2\pi = 1 \Leftrightarrow \pi = 0.50$$
$$x_P^E > x_D^E \Leftrightarrow \frac{A\alpha\pi}{1-\alpha} > \frac{A\alpha(1-\pi)}{1-\alpha} \Leftrightarrow \pi > 1-\pi \Leftrightarrow 2\pi > 1 \Leftrightarrow \pi > 0.50$$

**Observation 9.** Under both fee allocation rules with $\alpha \in [0,1)$, equilibrium trial expenditure increases (or remains constant) as the amount of the lawsuit increases:

$$\frac{\partial x_P^A}{\partial A} = \frac{\partial x_D^A}{\partial A} = \frac{\alpha}{4} \geq 0$$

$$\frac{\partial x_P^E}{\partial A} = \frac{\alpha\pi}{1-\alpha} \geq 0$$

$$\frac{\partial x_D^E}{\partial A} = \frac{\alpha(1-\pi)}{1-\alpha} \geq 0$$

**Observation 10.** Under both fee allocation rules, if the fixed costs are more than half the amount of the suit ($\frac{C}{A} > \frac{1}{2}$), no dispute will ever go to trial:

$$\begin{aligned}
\tfrac{C}{A} > \tfrac{1}{2} &\Rightarrow 4\tfrac{C}{A} > 2-\alpha \qquad \forall \alpha \in [0,1] \\
&\Rightarrow 8\tfrac{C}{A} > 4-2\alpha \\
&\Rightarrow 4\tfrac{C}{A} - \alpha > 4-3\alpha-4\tfrac{C}{A} \\
&\Rightarrow \frac{4\frac{C}{A}-\alpha}{4(1-\alpha)} > \frac{4-3\alpha-4\frac{C}{A}}{4(1-\alpha)}
\end{aligned}$$

Therefore, in this case, the conditions of Proposition 3 become

(i) Settlement $\Leftrightarrow \pi > \dfrac{4-3\alpha-4\frac{C}{A}}{4(1-\alpha)}$

(ii) No Suit $\Leftrightarrow \pi \leq \dfrac{4-3\alpha-4\frac{C}{A}}{4(1-\alpha)}$

(iii) Trial $\Leftrightarrow$ Otherwise

Thus, all disputes result in either settlement or no suit under the American rule.

$$\begin{aligned}
\tfrac{C}{A} > \tfrac{1}{2} &\Rightarrow \tfrac{C}{A} > \frac{1-2\alpha}{2(1-\alpha)} \qquad \forall \alpha \in [0,1] \\
&\Rightarrow 1-2\alpha < 2\tfrac{C}{A}(1-\alpha) \\
&\Rightarrow 1-\alpha < 2\tfrac{C}{A}(1-\alpha)+\alpha \\
&\Rightarrow \left(1-\tfrac{C}{A}\right)(1-\alpha) < \tfrac{C}{A}(1-\alpha)+\alpha
\end{aligned}$$

Therefore, in this case, the conditions of Proposition 4 become

(i) Settlement $\Leftrightarrow \pi > \left(1-\frac{C}{A}\right)(1-\alpha)$

(ii) No Suit $\Leftrightarrow \pi \leq \left(1-\frac{C}{A}\right)(1-\alpha)$

(iii) Trial $\Leftrightarrow$ Otherwise

Thus, all disputes result in either settlement or no suit under the English rule as well.

**Observation 11.** Under the American rule, if fixed costs are sufficiently small $\left(\frac{C}{A} < \frac{\alpha}{4}\right)$, all legal disputes will be resolved at trial:

If $\frac{C}{A} < \frac{\alpha}{4}$, the conditions of Proposition 3 become

(i) Settlement $\Leftrightarrow \pi > \dfrac{4-3\alpha-4\frac{C}{A}}{4-4\alpha} > \dfrac{4-4\alpha}{4-4\alpha} = 1$

(ii) No Suit $\Leftrightarrow \pi \leq \dfrac{4\frac{C}{A}-\alpha}{4(1-\alpha)} < \dfrac{\alpha-\alpha}{4(1-\alpha)} = 0$

(iii) Trial $\Leftrightarrow$ Otherwise

Since $0 \leq \pi \leq 1$, all disputes will be resolved at trial

**Observation 12.** If $\frac{C}{A} < \frac{1}{4}$, then every dispute that would go to trial under the English rule would also go to trial under the American rule:

First of all, it can be shown that

$$\frac{C}{A} < \frac{1}{4} \Rightarrow \left(1-\frac{C}{A}\right)(1-\alpha) < \frac{4-3\alpha-4\frac{C}{A}}{4(1-\alpha)}.$$

Thus, if the defendant prefers trial to settlement under the English rule, he will also prefer trial to settlement under the American rule. Furthermore, if we additionally note that we only need consider cases with $\alpha < \frac{1}{2}$ (we will show below that no cases go to trial under the English rule when $\alpha \geq \frac{1}{2}$), it can also be shown that

$$\tfrac{C}{A} < \tfrac{1}{4} \Rightarrow \tfrac{C}{A}(1-\alpha)+\alpha > \frac{4\frac{C}{A}-\alpha}{4(1-\alpha)}$$

Thus, if the plaintiff prefers trial to settlement under the English rule, he will also prefer trial to settlement under the American rule. Therefore, if $\frac{C}{A} < \frac{1}{4}$, then every dispute that would go to trial under the English rule would also go to trial under the American rule.

**Observation 13.** Under the English rule, a legal dispute will go to trial only if $\alpha < \pi \leq 1-\alpha$:

Suppose that $\pi > 1-\alpha$. In this case, we have $\pi > 1\text{-}\alpha \Rightarrow \pi > (1\text{-}\frac{C}{A})(1\text{-}\alpha)$, and such a legal dispute would therefore result in settlement under the English rule. Now suppose that a legal dispute does <u>not</u> result in settlement $\left(\text{i.e., } \pi \leq (1\text{-}\frac{C}{A})(1\text{-}\alpha)\right)$ and that $\pi \leq \alpha$. In this case, we have $\pi \leq \alpha \Rightarrow \pi \leq \frac{C}{A}(1\text{-}\alpha)+\alpha \Rightarrow \pi \leq \min\left\{(1-\frac{C}{A})(1\text{-}\alpha),\ \frac{C}{A}(1\text{-}\alpha)+\alpha\right\}$ and such a legal dispute would therefore result in no suit being filed under the English rule. Thus, a legal dispute will go to trial under the English rule only if $\alpha < \pi \leq 1-\alpha$.

Note that the above observation also implies that, if the productivity of lawyers is greater than or equal to one-half $(\alpha \geq \frac{1}{2})$, no dispute will ever go to trial under the English rule.

**Observation 14.** Under the American rule, if the fixed costs are less than one-fourth the amount of the suit $(\frac{C}{A} < \frac{1}{4})$, then the likelihood of trial increases (or does not change) as the productivity of lawyers, $\alpha$, increases:

$$\tfrac{C}{A} < \tfrac{1}{4} \Rightarrow \frac{\partial}{\partial\alpha}\left(\frac{4-3\alpha-4\frac{C}{A}}{4(1-\alpha)}\right) = \frac{1-4\frac{C}{A}}{16(1-\alpha)^2} > 0$$

$$\tfrac{C}{A} < \tfrac{1}{4} \Rightarrow \frac{\partial}{\partial\alpha}\left(\frac{4\frac{C}{A}-\alpha}{4(1-\alpha)}\right) = \frac{4\frac{C}{A}-1}{16(1-\alpha)^2} < 0$$

$$\begin{aligned}\tfrac{C}{A} < \tfrac{1}{4} &\Rightarrow 4\tfrac{C}{A} < 2-\alpha \qquad \forall\alpha\in[0,1]\\ &\Rightarrow 8\tfrac{C}{A} < 4-2\alpha\\ &\Rightarrow 4\tfrac{C}{A}-\alpha < 4-3\alpha-4\tfrac{C}{A}\\ &\Rightarrow \frac{4\frac{C}{A}-\alpha}{4(1-\alpha)} < \frac{4-3\alpha-4\frac{C}{A}}{4(1-\alpha)}\\ &\Rightarrow \min\left\{\frac{4-3\alpha-4\frac{C}{A}}{4(1-\alpha)}, \frac{4\frac{C}{A}-\alpha}{4(1-\alpha)}\right\} = \frac{4\frac{C}{A}-\alpha}{4(1-\alpha)}\end{aligned}$$

This means that the range of $\pi$ values for which settlement is predicted and the range of $\pi$ values for which no suit is predicted both get smaller as $\alpha$ increases. Therefore, if $\frac{C}{A} < \frac{1}{4}$, then as $\alpha$ increases, the likelihood of trial also increases.

**Observation 15.** Under the English rule, the likelihood of trial decreases (or does not change) as the productivity of lawyers, $\alpha$, increases:

Suppose that $\frac{C}{A}(1-\alpha)+\alpha \geq (1-\frac{C}{A})(1-\alpha)$. In this case, all disputes result in either settlement or no suit, so the likelihood of trial is zero for all $\alpha$. Now suppose instead that $\frac{C}{A}(1-\alpha)+\alpha < (1-\frac{C}{A})(1-\alpha)$. In this case, since $C < A$, we have that

$$\frac{\partial}{\partial\alpha}\left((1-\tfrac{C}{A})(1-\alpha)\right) = \tfrac{C}{A}-1 < 0$$

$$\frac{\partial}{\partial\alpha}\left(\tfrac{C}{A}(1-\alpha)+\alpha\right) = 1-\tfrac{C}{A} > 0.$$

This means that the range of $\pi$ values for which settlement is predicted and the range of $\pi$ values for which no suit is predicted both get larger as $\alpha$ increases. Therefore, the likelihood of trial decreases as $\alpha$ increases.

**Observation 16.** Under both fee allocation rules, as the merit of the case, $\pi$, increases, the likelihood of settlement increases and the likelihood of no suit decreases:

Under both rules, settlement occurs if $\pi$ is <u>greater</u> than some threshold while no suit occurs if $\pi$ is <u>less</u> than or equal to some other threshold. Therefore, as $\pi$ increases, the likelihood of settlement increases and the likelihood of no suit decreases.

**Observation 17.** Under both fee allocation rules, the likelihood of trial is greatest for closely contested lawsuits ($\pi$=0.50):

For any given value of $\pi$, the likelihood of trial depends upon the range of different $\alpha$ values for which trial is the predicted form of resolution. Under the American rule, trial occurs if and only if $\frac{4\frac{C}{A}-\alpha}{4(1-\alpha)} < \pi \leq \frac{4-3\alpha-4\frac{C}{A}}{4(1-\alpha)}$. The likelihood of trial is therefore maximized when $\pi$ is precisely the midpoint between the lower and upper bounds of this inequality. This midpoint is given by:

$$\frac{1}{2}\left(\frac{4\frac{C}{A}-\alpha}{4(1-\alpha)}+\frac{4-3\alpha-4\frac{C}{A}}{4(1-\alpha)}\right) = \frac{4\frac{C}{A}-\alpha+4-3\alpha-4\frac{C}{A}}{8(1-\alpha)} = \frac{4-4\alpha}{8(1-\alpha)} = \frac{1}{2}.$$

Under the English rule, trial occurs if and only if $\frac{C}{A}(1-\alpha)+\alpha < \pi \leq \left(1-\frac{C}{A}\right)(1-\alpha)$. The likelihood of trial is again maximized when $\pi$ is equal to the midpoint between the lower and upper bounds of this inequality. This midpoint is given by:

$$\tfrac{1}{2}\left(\tfrac{C}{A}(1-\alpha)+\alpha+\left(1-\tfrac{C}{A}\right)(1-\alpha)\right) = \tfrac{1}{2}\left(\alpha+(1-\alpha)\right) = \tfrac{1}{2}.$$

Thus, under both fee allocation rules, the likelihood of trial is highest when $\pi$=0.50.

**Observation 18.** There exist additional Nash equilibria which are not subgame perfect. These Nash equilibria are characterized by strategies off the equilibrium path in which the defendant chooses to go to trial when he would prefer settlement or in which one party chooses a very high level of legal expenditure at trial making trial prohibitively unattractive to the other party.

## 5 EXPERIMENTAL RESULTS

The experimental results under the different parameter configurations are summarized in Figure 7 for the American rule and Figure 8 for the English rule. For each fee allocation rule, 320 experimental legal disputes were conducted, and therefore the behavior of 640 litigants was observed. Not included in these numbers and not reflected in Figures 7 and 8 are the experimental legal disputes that were conducted under the uncertain merit conditions. The uncertain merit experiments account for 340 additional disputes and will be discussed separately below.

In this section we discuss the patterns of subject behavior observed in the experimental sessions and discuss the influence of various factors on this behavior. These experimental results are broken into five subject areas: (1) behavior under alternative allocation rules, (2) impact of lawyer productivity, (3) influence of case merit, (4) effect of uncertain merit, and (5) performance of model predictions.

## 5.1 BEHAVIOR UNDER ALTERNATIVE ALLOCATION RULES

The first three results summarize litigant behavior under the two legal fee allocation rules. If a dispute is resolved at trial, the total legal expenditure at trial is greater under English rule than under American rule (Result 1). While the English rule does discourage trials (Result 2) this effect is not strong enough to offset the greater expenditure. The net effect of a move to the English rule is to increase legal expenditure per dispute (Result 3).

**Result 1.** The English rule produces significantly greater legal expenditure at trial than the American rule.

**Support.** Figure 9 shows that 96% of all trial expenditures under the American rule were at or below 100 francs, 100% were at or below 200 francs, and the mean expenditure was 45 francs. On the other hand, trial expenditures under the English rule were distributed throughout the allowed range of 0 to 1000 francs with a mean expenditure of 580 francs, almost 13 times higher than the mean under the American rule. The difference in mean expenditure under the two different rules is statistically significant at the 1% level. Furthermore, Figures 7 and 8 indicate that for every one of the nine combinations of $\alpha$ and $\pi$ and for both plaintiff and defendant, mean expenditure at trial was always at least 5.3 times larger under the English rule than under the American rule ($\alpha$=0.25, $\pi$=0.75, defendant) and was as much as 28.6 times larger ($\alpha$=0.50, $\pi$=0.75, plaintiff).

**Result 2.** Under the English rule, legal disputes are less likely to result in a trial than under the American rule.

**Support.** Figure 10 shows that 80% of all disputes were resolved at trial under the American rule while only 12% of all disputes were resolved at trial under the English rule. This difference in proportion of disputes resolved at trial under the two different rules is statistically significant at the 1% level. Furthermore, Figures 7 and 8 reveal that for every one of the nine combinations of $\alpha$ and $\pi$, the frequency of trial was always at least 2.8 times higher under the American rule than under the English rule ($\alpha$=0.25, $\pi$=0.50). For all nine parameter combinations, no fewer than 60% ($\alpha$=0.50, $\pi$=0.25) and as many as 96% ($\alpha$=0.50, $\pi$=0.50) of all disputes were resolved at trial under the American rule. In contrast, no more than 34% ($\alpha$=0.25, $\pi$=0.50) and as few as 0% ($\alpha$=0.50, $\pi$=0.25, and $\alpha$=0.75, $\pi$=0.75) of disputes resulted in trial under the English rule.

**Result 3.** Total expenditure per legal dispute is higher under the English rule than under the American rule

**Support.** According to the data from Figures 9 and 10, under the American rule, trial occurred in 80.0% of all disputes and the mean expenditure at trial was 44.9 francs. Thus, the average expenditure at trial per person per dispute under the American rule was 35.9 francs. If we also include the fixed costs incurred for cases that were resolved by settlement or trial, this figure becomes 44.3 francs. Under the English rule, trial occurred in 12.5% of the cases and the mean expenditure at trial was 580.2 francs. Thus, the average expenditure at trial per person per dispute under the English rule was 75.5 francs. If we also include the fixed costs incurred for cases that were resolved by settlement or trial, this figure becomes 78.7 francs, approximately 78% higher than under the American rule. This difference in mean expenditure per dispute under the two different rules, with or without inclusion of the fixed costs, is statistically significant at the 1% level.

The next three subsections explore several parameters that influence the level of expenditure and form of resolution in a legal dispute. The dispute parameters investigated are lawyer productivity, case merit, and uncertainty of merit. After the impact of these factors is discussed, the analysis moves in subsection 4.5 to consider models that may serve as underlying explanations of the effects of different allocation rules and dispute parameters.

## 5.2 IMPACT OF LAWYER PRODUCTIVITY

**Result 5.** Under both fee allocation rules, legal expenditure at trial increases as the productivity of lawyers increases. This trend is more significant under the English rule than under the American rule.

**Support.** Figure 11 clearly illustrates that mean legal expenditure at trial is higher for higher values of $\alpha$ under both the American and English rules. This trend is particularly significant under the English rule with mean expenditure jumping from 438 when $\alpha$=0.25 to 630 when $\alpha$=0.50 to the expenditure ceiling of 1000 when $\alpha$=0.75. The difference in mean expenditure between $\alpha$=0.25 and $\alpha$=0.50 under the English Rule is statistically significant at the 2% level while the difference in mean expenditure between $\alpha$=0.50 and $\alpha$=0.75 is statistically significant at the 1% level. Mean expenditure under the American rule, on the other hand, increases more modestly from 35 to 49 to 53 for the three different levels of $\alpha$. The difference in mean expenditure between $\alpha$=0.25 and $\alpha$=0.50 under the American Rule is statistically significant at the 1% level; however, the difference in mean expenditure between $\alpha$=0.50 and $\alpha$=0.75 is not statistically significant.

**Result 5.** Under both fee allocation rules, the frequency of trial decreases as the productivity of lawyers increases. This trend is more significant under the English rule than under the American rule.

**Support.** Figure 12 illustrates the frequency of trial for various levels of $\alpha$. As $\alpha$ changes from 0.25 to 0.50 to 0.75, the percentage of disputes resolved at trial under the American rule drops from 83% to 81% to 76%; however, neither of these differences in percentages are statistically significant. Similarly, the percentage of disputes resolved at trial under the English rule drops from 17% to 15% to 5%. The latter difference (between $\alpha$=0.50 and $\alpha$=0.75) is statistically significant at the 1% level in this case.

## 5.3 INFLUENCE OF CASE MERIT

**Result 6.** Under both fee allocation rules, defendant expenditure at trial exceeds plaintiff expenditure at trial for frivolous lawsuits ($\pi$=0.25) while plaintiff expenditure at trial exceeds defendant expenditure at trial for strong lawsuits ($\pi$=0.75). The expenditures at trial for the two parties are most similar for closely contested lawsuits ($\pi$=0.50).

**Support.** Figure 13 demonstrates that under both allocation rules, mean defendant expenditure at trial is higher than mean plaintiff expenditure at trial when $\pi$=0.25, while the opposite relationship is true when $\pi$=0.75. These differences in expenditure are most significant for $\pi$=0.75 (at the 5% level under the American Rule and at the 12% level under the English Rule). The difference between mean plaintiff and mean defendant expenditure reaches a minimum of 2 under the American rule and a minimum of 72 under the English rule, both at $\pi$=0.50.

**Result 7.** Under both fee allocation rules, frivolous lawsuits ($\pi$=0.25) are the most likely to not be filed, closely contested lawsuits ($\pi$ =0.50) are the most likely to be resolved at trial, and strong lawsuits ($\pi$ =0.75) are the most likely to produce a pretrial settlement.

**Support.** Figure 14 illustrates the frequency of the forms of resolution for various levels of $\pi$. Under the American rule, the frequency of no suit reaches a peak of 36% for $\pi$=0.25, the frequency of trial reaches a peak of 89% for $\pi$=0.50, and the frequency of settlement reaches a peak of 16% for $\pi$=0.75. Similarly, under the English rule, the frequency of no suit reaches a peak of 93% for $\pi$=0.25, the frequency of trial reaches a peak of 23% for $\pi$=0.50, and the frequency of settlement reaches a peak of 88% for $\pi$=0.75. The differences between the peak percentage and the other percentages for each form of resolution is statistically significant at the 1% level in all but two cases and at the 5% level in all but one case (percentage of trials under the American Rule between $\pi$=0.50 and $\pi$=0.75).

## 5.4 EFFECT OF UNCERTAIN MERIT

**Result 8.** Under both fee allocation rules, legal expenditure at trial is lower when the merit of the lawsuit is uncertain than when the merit is known.

**Support.** Figure 15 indicates that under the American rule, mean expenditure at trial drops from 44.9 to 36.5 with the addition of uncertain merit. Figure 16 indicates that under the English rule, mean expenditure at trial drops from 580.2 to 439.6 with the addition of uncertain merit.

**Result 9.** Under both fee allocation rules, the frequency of trial is higher when the merit of the lawsuit is uncertain than when the merit is known.

**Support.** Figure 15 illustrates that under the American rule, the frequency of trial increases from 80% to 85% with the addition of uncertain merit. Figure 16 illustrates that under the English rule, the frequency of trial increases more than two-fold from 12% to 26% with the addition of uncertain merit.

**Result 10.** The difference in expenditure per dispute between the American and English rules is greater when the merit of the lawsuit is uncertain than when the merit is known.

**Support.** Calculating expenditure per person per dispute as before (see Result 3), we discover that under the American rule, average expenditure per person per dispute <u>decreases</u> from 35.92 to 31.13 (44.31 to 39.94 including fixed costs) with the addition of uncertain merit. On the other hand, under the English rule, average expenditure per person per dispute <u>increases</u> from 75.53 to 113.86 (78.66 to 118.23 including fixed costs) with the addition of uncertain merit. Thus the difference in expenditure per dispute between the two rules increases with the addition of uncertainty.

## 5.5 PERFORMANCE OF MODEL PREDICTIONS

While the general parametric influence on legal expenditure and dispute resolution is of great interest, it is also important to explore why these factors have the influence that they do. In particular it is important to inquire about the reliability of game theoretic models in helping us understand the patterns of data. Where are they accurate and where do they tend to fail?

The first several results in this section (Result 11 through Result 17) tell us that the qualitative predictions of the Nash equilibrium and subgame perfect equilibrium models are almost always consistent with the observed experimental behavior and outcomes. These results suggest that traditional game theory contributes significantly to our understanding of the relative institutional response to changes in fee allocation rule, case merit, and lawyer productivity. On the other hand, the latter results of this section (Result 18 through Result 20) identify certain areas in which the specific quantitative predictions of the game theoretic models are inconsistent with the experimental observations.

**Result 11.** The direction of the difference in expenditure at trial under the two different allocation rules is as predicted by the Nash equilibrium model.

**Support.** Observation 1 indicates that the Nash equilibrium model predicts, for all experimental parameters, that legal expenditure at trial will be higher under the English rule than under the American rule. This prediction matches Result 1 presented above. In addition, Observation 1 specifically says that total expenditure at trial should always be at least twice as large under the English rule as under the American rule. Comparison between Figures 7 and 8 indicates that this is true for all combinations of $\alpha$ and $\pi$.

**Result 12.** The direction of the difference in frequency of trial under the two different allocation rules is as predicted by the subgame perfect equilibrium model.

**Support.** For the parameters used in the experimental sessions (A=240, C=10), Observation 12 says that the subgame perfect equilibrium model predicts that the frequency of trial will be lower under the English rule than under the American rule. This prediction matches Result 2 presented above.

**Result 13.** For almost all parameter combinations, the most frequently observed form of resolution is the form predicted by the subgame perfect equilibrium model.

**Support.** Figure 5 illustrates the form of resolution predicted by the subgame perfect equilibrium model under the American rule for the particular values used in the experimental sessions (A=240, C=10). This figure shows that trial is the predicted form of resolution under the American rule for all nine combinations of $\alpha$ and $\pi$ used in the experiments. Comparing this prediction with the experimental results in Figure 7 reveals that trial is, in fact, the most frequently observed form of resolution under the American rule for all parameter combinations. For the English rule, the crosses in Figure 6 illustrate the forms of resolution predicted by the subgame perfect equilibrium model for the nine combinations of $\alpha$ and $\pi$ used in the experiments. Comparing these predictions with the experimental results in Figure 8 reveals that the most frequently observed form of resolution matches the predicted form of resolution for seven of the nine combinations of $\alpha$ and $\pi$. Combining the results from both rules, the most frequently observed form of resolution matches the predicted form of resolution in 16 out of the 18 different parameter combinations (three levels of $\alpha$, three levels of $\pi$, and two different allocation rules).

**Result 14.** Under both fee allocation rules, the effect of changes in the productivity of lawyers on legal expenditure at trial is as predicted by the Nash equilibrium model.

**Support.** Under both fee allocation rules, Observation 3 says that the Nash equilibrium model predicts that legal expenditure at trial will increase as the productivity of lawyers increases. This prediction matches Result 4 presented above. Moreover, Figures 3 and 4 demonstrate that the Nash model predicts that the increase in legal expenditure as a response to an increase in lawyer productivity will be more significant under the English rule than under the American rule. This prediction is also verified by Result 4 above.

**Result 15.** Under the English rule, the effect of changes in the productivity of lawyers on the frequency trial is as predicted by the subgame perfect equilibrium model.

**Support.** Observation 15 indicates that, under the English rule, the subgame perfect equilibrium model predicts that the frequency of trial will decrease as the productivity of lawyers increases. This prediction coincides with Result 5 presented above.

**Result 16.** Under the English rule, the effect of changes in case merit on legal expenditure at trial is as predicted by the Nash equilibrium model.

**Support.** According to Observations 7 and 8, the Nash equilibrium model predicts that (a) defendant expenditure at trial will exceed plaintiff expenditure at trial when $\pi$=0.25, (b) plaintiff expenditure at trial will exceed defendant expenditure at trial when

$\pi$=0.75, and (c) the difference between plaintiff and defendant expenditure at trial should be smallest for $\pi$=0.50. All three of these predictions are verified by Result 6 above.

**Result 17.** Under both fee allocation rules, the effect of changes in case merit on the frequency of suit, settlement, and trial is as predicted by the subgame perfect equilibrium model.

**Support.** According to Observations 16 and 17, the subgame perfect equilibrium model predicts that, under both allocation rules, frivolous lawsuits ($\pi$=0.25) will be the most likely to not be filed, closely contested lawsuits ($\pi$=0.50) will be the most likely to be resolved at trial, and strong lawsuits ($\pi$=0.75) will be the most likely to produce a pretrial settlement. This prediction coincides with Result 7 presented above.

**Result 18.** Under the American rule, average legal expenditure at trial is slightly higher than predicted by the Nash equilibrium model. Under the English rule, average legal expenditure at trial is much higher than predicted by the Nash equilibrium model.

**Support.** Figure 17 shows that for all values of $\alpha$, the observed average expenditure at trial under both allocation rules is above the level of expenditure predicted by the Nash equilibrium model. This figure also illustrates that the difference between observed and predicted expenditure at trial is much more significant under the English rule than under the American rule (note the different scales for the vertical axes in the figure). In addition, comparison of predicted expenditure levels in Figures 3 and 4 to observed expenditure levels in Figures 7 and 8 allow examination of differences for all nine combinations of $\alpha$ and $\pi$. Under the American rule, observed expenditure at trial ranges from 10% <u>below</u> prediction ($\alpha$=0.50, $\pi$=0.75, defendant) to 220% above prediction

($\alpha$=0.25, $\pi$=0.50, plaintiff). Under the English rule, observed expenditure at trial ranges from 85% above prediction ($\alpha$=0.75, $\pi$=0.25, defendant) to 1415% above prediction ($\alpha$=0.25, $\pi$=0.50, defendant). All differences between observed and predicted expenditure levels are statistically significant at the 1% level.

**Result 19.** Under the American rule, the frequency of trial is lower than predicted by the subgame perfect equilibrium model.

**Support.** Figure 18 illustrates that the subgame perfect equilibrium model predicts 100% of legal disputes will go to trial under the American rule for the particular parameter values used in the experimental sessions. This figure also shows, however, that only 80% of all experimental disputes are actually resolved at trial. Moreover, Figure 7 indicates that, for particular combinations of $\alpha$ and $\pi$, as few as 60% of disputes are resolved at trial under the American rule.

**Result 20.** Under the English rule, the frequency of no suit is higher than predicted by the subgame perfect equilibrium model while the frequency of settlement is lower than predicted.

**Support.** As illustrated in Figure 19, the subgame perfect equilibrium model predicts that, under the English rule, 21% of all disputes will result in no suit being filed, 67% will result in pretrial settlement, and 12% will proceed to trial (note that these percentages are determined by the observed relative frequency of the different combinations of $\alpha$ and $\pi$ in the experimental sessions). Figure 19 also depicts the observed frequency of the different forms of resolution, and although the observed frequency of trial (12%) matches the prediction, the observed frequency of no suit (50%) is significantly greater than predicted while the observed frequency of settlement (38%) is significantly lower than predicted. Comparing predictions and observations for specific parameter values reveals that much of the overall discrepancy can be traced to two specific parameter combinations: $\alpha$=0.75, $\pi$=0.25, and $\alpha$=0.75, $\pi$=0.50. Figure 6 illustrates that settlement is the predicted form of resolution under both of these parameter combinations; however, in both cases, Figure 8 reveals that the most frequently observed resolution is no suit being filed (90% and 67% of disputes), with settlement occurring much less frequently (0% and 37% of disputes). Note that both litigants prefer to avoid trial under these parameter combinations; however, the subgame perfect equilibrium model predicts that the plaintiff will file suit with the knowledge (or belief) that the defendant will subsequently choose to settle rather than go to trial. In the actual experiments, however, many plaintiffs are choosing not to file suit, apparently because they fear that the defendants will "call their bluff" and proceed to trial.

# 6 EX-POST THEORIZING AND CONJECTURES

The analysis in this paper provides important insight into the impact of alternative legal fee allocation rules on the behavior of litigants and the resolution of legal disputes. Nonetheless, there remain relevant unanswered questions and significant avenues for further research in the field. In this section, we present rudimentary theories on several issues that are raised or unaddressed by our analysis and discuss potential research extensions that are outside the scope of the present paper.

As mentioned previously, a comprehensive investigation of different fee allocation rules requires examination of all four stages in the chronology of a legal dispute (Figure 1), recognizing that behavior in each preliminary stage will depend heavily upon expectations about the outcome of later stages. The present paper is intended to be a first step in such an investigation, and therefore focuses primarily on the different effects of the American and English rules on outcomes and decisions at trial, the final stage in the chronology. Other researchers may seek to extend our analysis to the previous stage of settlement bargaining, and in doing so may employ a more flexible settlement procedure than the strict forfeiture settlement mechanism used in our investigation. It is therefore sensible to discuss the anticipated effects of alternative settlement mechanisms on the results of this paper.

It is reasonable to expect that a more flexible settlement mechanism could produce additional settlements and fewer trials than were predicted and observed in the present analysis. Recall that the forfeiture settlement mechanism we employed was chosen with the expectation that the number of disputes resolved at trial would be significant enough for us to draw strong conclusions about trial expenditure decisions. In our experiments, more than 90% of the lawsuits filed under the American rule were resolved at trial, whereas fewer than 10% of all lawsuits proceed to trial in actual practice. Therefore, any settlement mechanism that is selected to more closely represent existing legal procedure should result in a greater number of lawsuits being settled out of court.

Despite the prospect of increasing the settlement rate, use of an alternative mechanism is nevertheless unlikely to reverse any of the results comparing litigant behavior under the two different allocation rules. For example, trial expenditure should continue to be higher under the English rule than under the American rule (Result 1), since the settlement mechanism has no effect on incentives at trial (although it may influence the type of disputes that proceed to trial). Moreover, as long as this disparity in trial expenditure persists, there will be a greater incentive to settle and therefore a lower frequency of trial under the English rule than under the American rule (Result 2).

We also expect that expenditure per legal dispute would continue to be higher under the English rule than under the American rule (Result 3) for any reasonable settlement mechanism. Given that the observed expenditure per trial under the English rule was 13 times higher than under the American rule, adoption of an alternative settlement mechanism would reverse our result only if the new mechanism produced 13 trials under the American rule for every single trial under the English rule. No matter the settlement procedure, such a significant difference in trial rates is highly unlikely and inconsistent with empirical evidence [Hughes and Snyder 1991, 1995].

Another potential research extension is the enhancement of the game theoretic models to explain the discrepancies between predicted behavior and experimental observations. The models presented in this paper are remarkably effective in terms of predicting the qualitative behavioral impact of changes in fee allocation rule, case merit, and lawyer productivity. Nonetheless, there are experimental treatments in which the observed form of resolution and/or level of legal expenditure differs significantly from the model predictions.

First of all, it is possible that a model of litigant behavior containing an element of randomness or imperfect performance may explain some of the observed actions and outcomes that are inconsistent with the traditional game theoretic model. Introducing small errors in performance could account for cases in which observed litigant behavior does not differ substantially from the prediction. For example, although the observed form of resolution is most frequently the form predicted, we still observe dispute resolutions that are zero likelihood events according to the model. With the introduction of randomness or error, this zero likelihood problem is immediately averted as all possible outcomes become positive probability events.

In addition, the results of the uncertain merit experiments suggest that a model incorporating uncertain or asymmetric information may also have considerable explanatory power. As previously discussed, the addition of uncertain merit increases the frequency of trial, especially under the English rule. Information uncertainty therefore presents itself as a potential explanation for the occurrence of trials (with frequencies as high as 28%) under the English rule in treatments for which settlement or no suit is the predicted resolution. In particular, a litigant may be uncertain about the interpretation of the dispute process, about the assessment of the probability of prevailing at trial, or about the opposing parties beliefs about these same factors.

Lastly, anomalous litigant behavior may also be a result of non-neutral attitudes toward risk. In particular, consider the most dramatic inaccuracy of the current game theoretic model, which is the significant underestimation of legal expenditure at trial under the English rule. In these disputes, the equilibrium expected profit at trial is always negative for the defendant and is negative for the plaintiff in six of nine treatments. Since trial is a negative value gamble for both parties in such cases, prospect theory [Kahneman and Tversky 1979] predicts risk seeking behavior by the litigants, making them more inclined to take the greater gamble associated with larger trial expenditures. Such risk attitudes may therefore explain why observed expenditure at trial under the English rule is significantly higher than the current model predicts for risk neutral parties. In addition, it is possible that non-neutral risk attitudes may be the rationale for other observed behavior that is inconsistent with the game theoretic models as currently constructed.

# 7 . CONCLUSIONS

The analysis of legal fee allocation rules presented in this paper suggests that a change from the American rule to the English rule could result in extreme changes in the legal process. The experimental results as well as the game theoretic model applied to the legal dispute environment under investigation indicate significant differences in the level of legal expenditures and the frequency of suit, settlement, and trial induced by the two rules.

In the experimental legal environment, subjects chose levels of expenditure at trial under the English rule which were on average almost 13 times larger than the levels of expenditure at trial chosen under the American rule. On the other hand, nearly 6 times fewer legal disputes were brought to trial under the English rule than under the American rule. Despite the lower frequency of trial under the English rule, total expenditure per dispute was 78% higher under the English rule than under the American rule.

These results indicate that while a move to the English rule may reduce the number of lawsuits and trials in our legal system, it may nevertheless increase the total cost of the system as a result of dramatically increased expenditure at trial. The surprisingly high legal fees that must be paid by a losing party under the English rule also raises significant issues concerning proper access to justice. Parties with meritorious claims may be deterred from going to trial or even using the legal system at all when the potential costs are so high. It is a fundamental premise of our legal system that every citizen is entitled to her day in court, and relieving court congestion may not be justified if, as a consequence, potential litigants are afraid to exercise their legal rights.

In addition to the qualitative differences between the American rule and English rule, we were also able to identify the impact of several other factors on litigation expenditure and dispute resolution. The productivity of lawyers was shown to be positively related to legal expenditure at trial and negatively related to the frequency of trial. Case merit was also found to have significant effects, with frivolous lawsuits being the most likely to not be filed, closely contested lawsuits the most likely to be resolved at trial, and strong lawsuits the most likely to produce a pretrial settlement. In addition, defendants outspent plaintiffs on average when frivolous lawsuits were resolved at trial while plaintiffs outspent defendants on average when strong lawsuits were resolved at trial. Finally, the effect of uncertain merit was to decrease expenditure at trial, increase the frequency of trial, and increase the gap between the American rule and English rule in terms of expenditure per dispute.

The Nash equilibrium and subgame perfect equilibrium models provide accurate predictions regarding the qualitative differences between the American and English rules as well as the impact of changes in lawyer productivity and case merit. Nonetheless, the specific quantitative predictions were not always accurate, with the most dramatic discrepancy being a significant underestimation of the level of legal expenditure at trial under the English rule. Directions for future research include enhancements to the current models that may explain such discrepancies, perhaps incorporating errors in performance, uncertain or asymmetric information, or non-neutral attitudes toward risk.

## Figure 1

## Chronology of a Legal Dispute

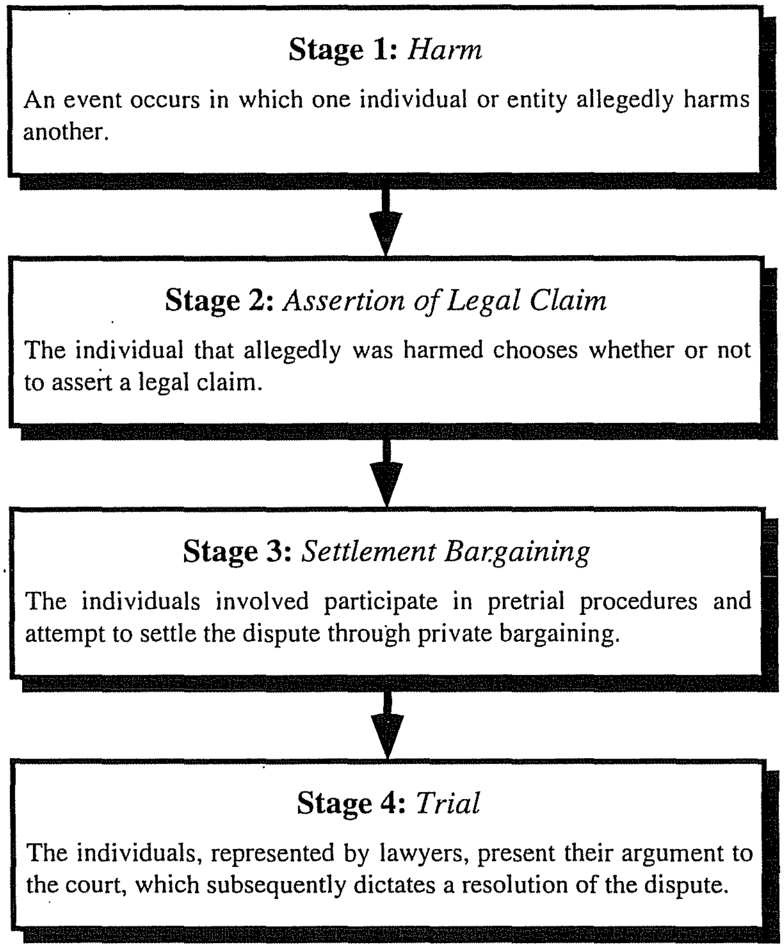


**Figure 2**

**Structure of Experimental Legal Dispute**

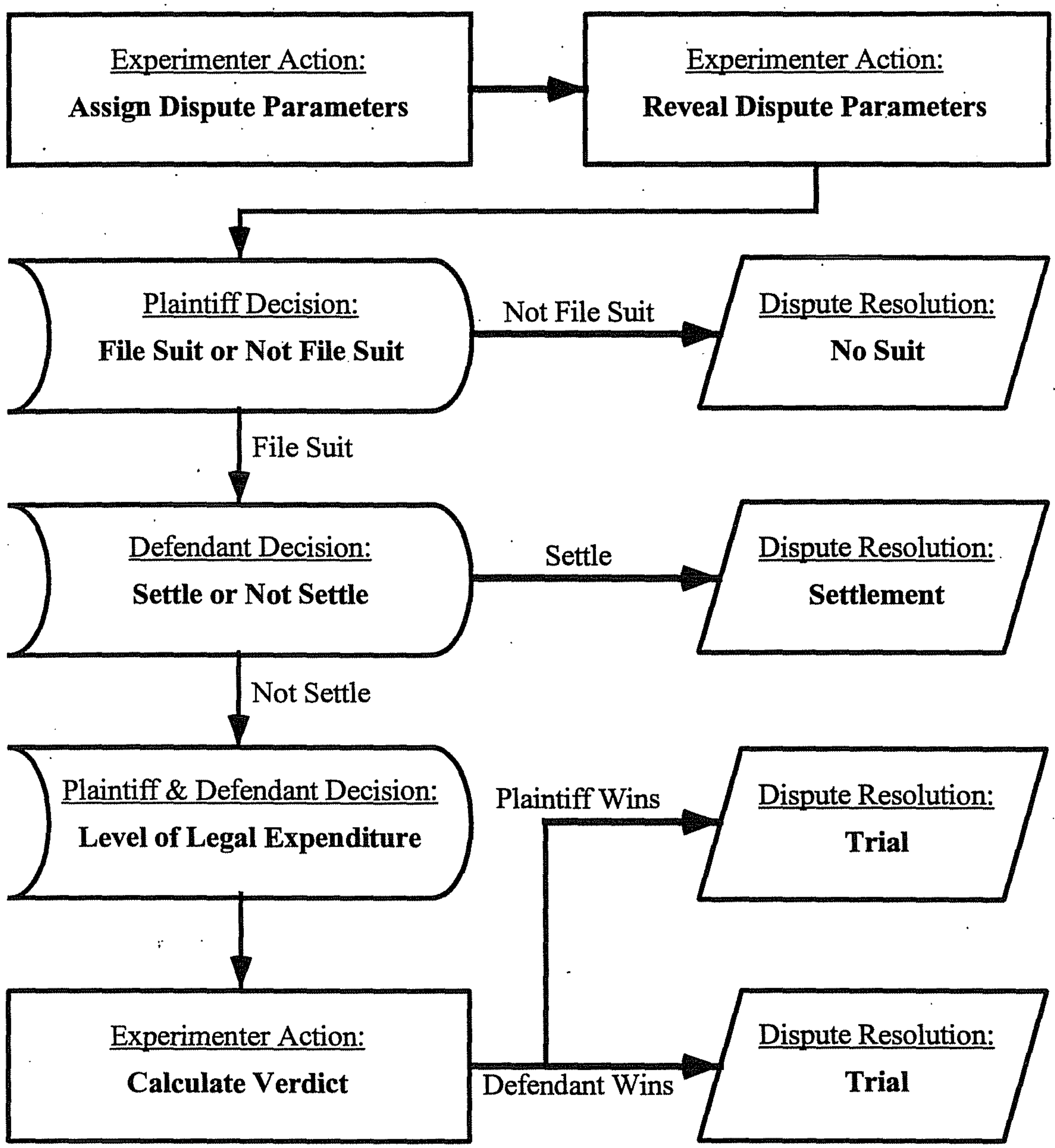


**Figure 3**

**Predicted Expenditure At Trial Under American Rule (A=240, C=10)**

| | | | π | | |
|---|---|---|---|---|---|
| | | | 0.25 | 0.50 | 0.75 |
| | 0.25 | PLAINTIFF: | 15 | 15 | 15 |
| | | DEFENDANT: | 15 | 15 | 15 |
| | | TOTAL | 30 | 30 | 30 |
| α | 0.50 | PLAINTIFF: | 30 | 30 | 30 |
| | | DEFENDANT: | 30 | 30 | 30 |
| | | TOTAL | 60 | 60 | 60 |
| | 0.75 | PLAINTIFF: | 45 | 45 | 45 |
| | | DEFENDANT: | 45 | 45 | 45 |
| | | TOTAL | 90 | 90 | 90 |

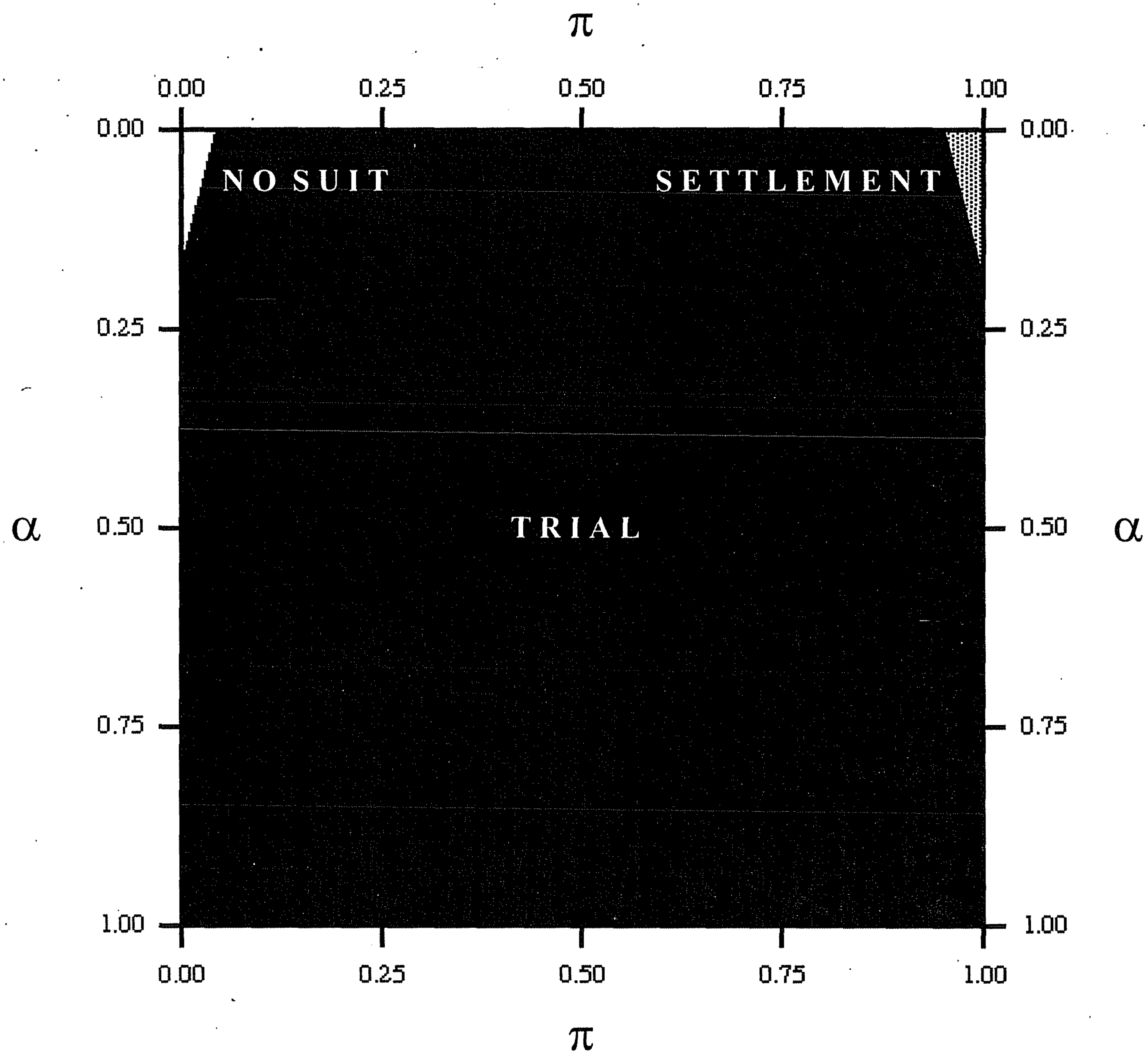
**Figure 4**

**Predicted Expenditure At Trial Under English Rule (A=240, C=10)**

| | | | π | |
|---|---|---|---|---|
| | | 0.25 | 0.50 | 0.75 |
| α 0.25 | PLAINTIFF: | 20 | 40 | 60 |
| | DEFENDANT: | 60 | 40 | 20 |
| | TOTAL | 80 | 80 | 80 |
| α 0.50 | PLAINTIFF: | 60 | 120 | 180 |
| | DEFENDANT: | 180 | 120 | 60 |
| | TOTAL | 240 | 240 | 240 |
| α 0.75 | PLAINTIFF: | 180 | 360 | 540 |
| | DEFENDANT: | 540 | 360 | 180 |
| | TOTAL | 720 | 720 | 720 |

**Figure 5**

**Predicted Form of Resolution Under American Rule (A=240, C=10)**



**<u>Figure 6</u>**

**Predicted Form of Resolution Under English Rule (A=240, C=10)**

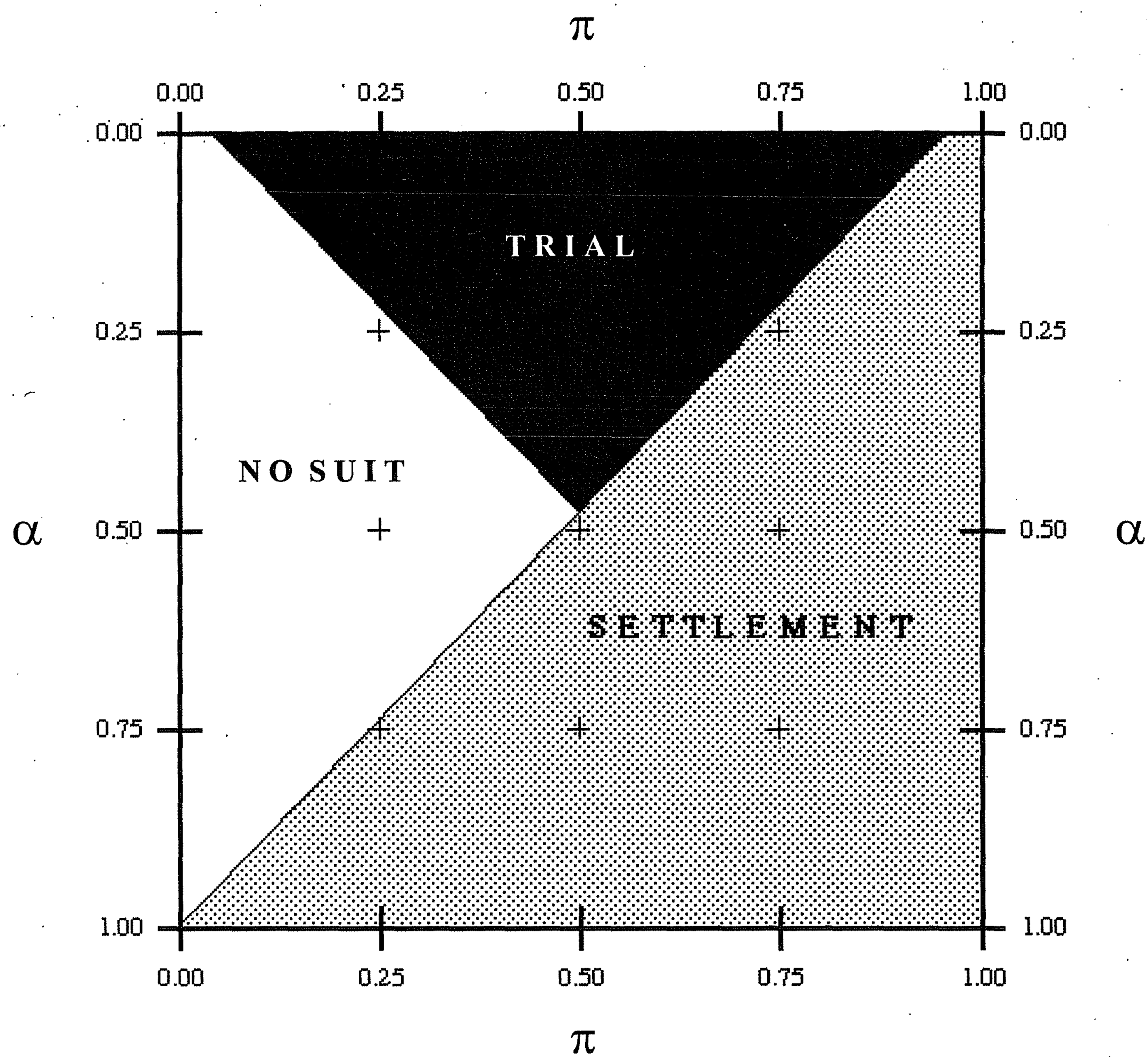


**Figure 7**

**Experimental Results Under American Rule**

| | | | π | |
|---|---|---|---|---|
| α | | 0.25 | 0.50 | 0.75 |
| 0.25 | MEAN EXPENDITURE | | | |
| | PLAINTIFF | 27 | 48 | 32 |
| | DEFENDANT | 37 | 38 | 23 |
| | FORM OF RESOLUTION | | | |
| | NO SUIT | 39% | 5% | 0% |
| | SETTLEMENT | 0% | 0% | 10% |
| | TRIAL | 61% | 95% | 90% |
| 0.50 | MEAN EXPENDITURE | | | |
| | PLAINTIFF | 44 | 51 | 35 |
| | DEFENDANT | 60 | 72 | 27 |
| | FORM OF RESOLUTION | | | |
| | NO SUIT | 40% | 2% | 0% |
| | SETTLEMENT | 0% | 2% | 21% |
| | TRIAL | 60% | 96% | 79% |
| 0.75 | MEAN EXPENDITURE | | | |
| | PLAINTIFF | 54 | 62 | 55 |
| | DEFENDANT | 44 | 55 | 49 |
| | FORM OF RESOLUTION | | | |
| | NO SUIT | 26% | 12% | 0% |
| | SETTLEMENT | 4% | 14% | 18% |
| | TRIAL | 70% | 74% | 82% |

**Figure 8**

**Experimental Results Under English Rule**

| α | | π = 0.25 | π = 0.50 | π = 0.75 |
|---|---|---|---|---|
| 0.25 | MEAN EXPENDITURE | | | |
| | PLAINTIFF | 295 | 378 | 377 |
| | DEFENDANT | 510 | 606 | 123 |
| | FORM OF RESOLUTION | | | |
| | NO SUIT | 90% | 53% | 10% |
| | SETTLEMENT | 0% | 13% | 83% |
| | TRIAL | 10% | 34% | 7% |
| 0.50 | MEAN EXPENDITURE | | | |
| | PLAINTIFF | - | 635 | 1000 |
| | DEFENDANT | - | 557 | 700 |
| | FORM OF RESOLUTION | | | |
| | NO SUIT | 100% | 32% | 4% |
| | SETTLEMENT | 0% | 40% | 89% |
| | TRIAL | 0% | 28% | 7% |
| 0.75 | MEAN EXPENDITURE | | | |
| | PLAINTIFF | 1000 | 1000 | - |
| | DEFENDANT | 1000 | 1000 | - |
| | FORM OF RESOLUTION | | | |
| | NO SUIT | 90% | 60% | 4% |
| | SETTLEMENT | 0% | 37% | 96% |
| | TRIAL | 10% | 3% | 0% |

**Figure 9**

**Expenditure at Trial Under Alternative Allocation Rules**

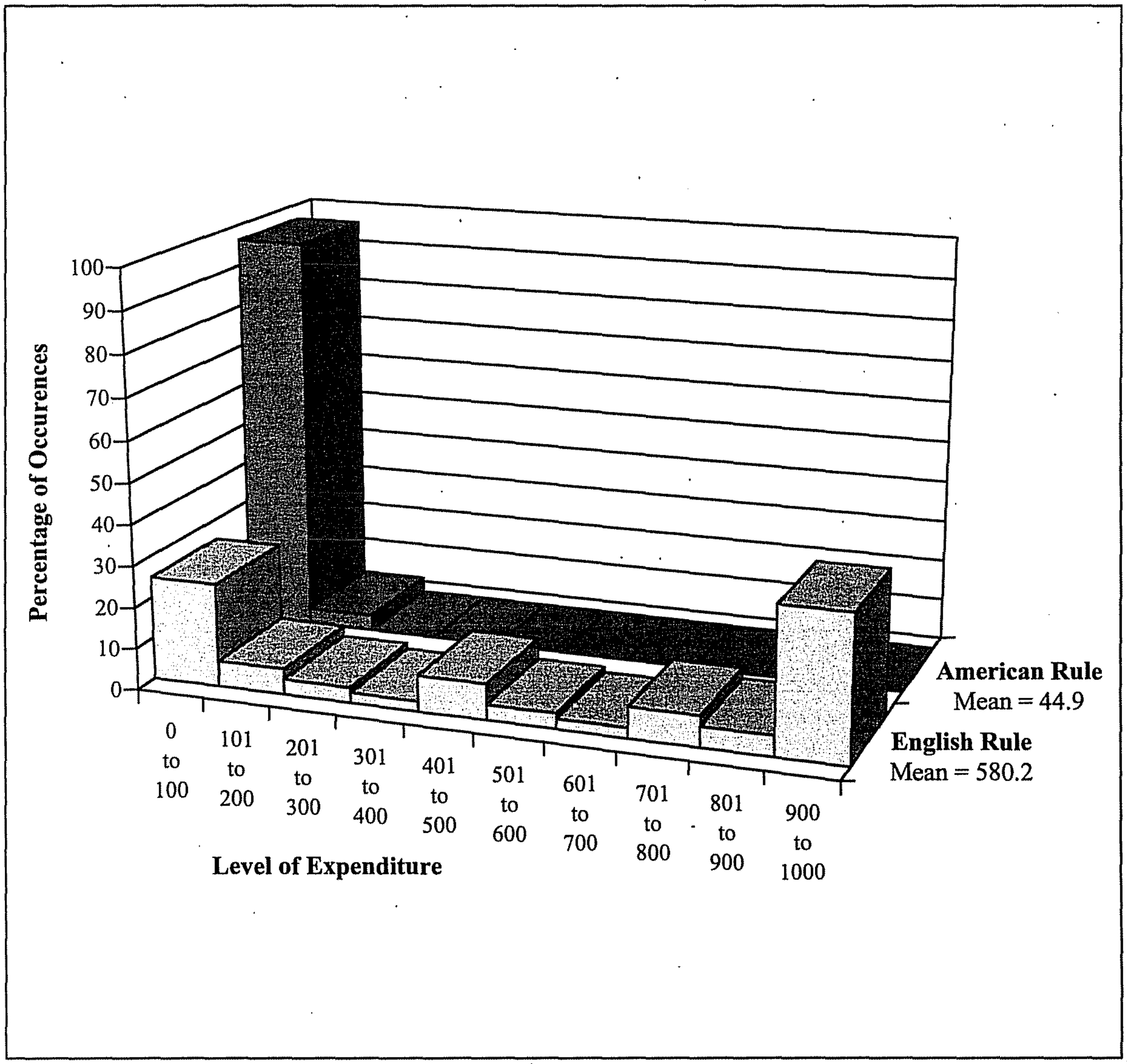


**Figure 10**

**Form of Resolution Under Alternative Allocation Rules**

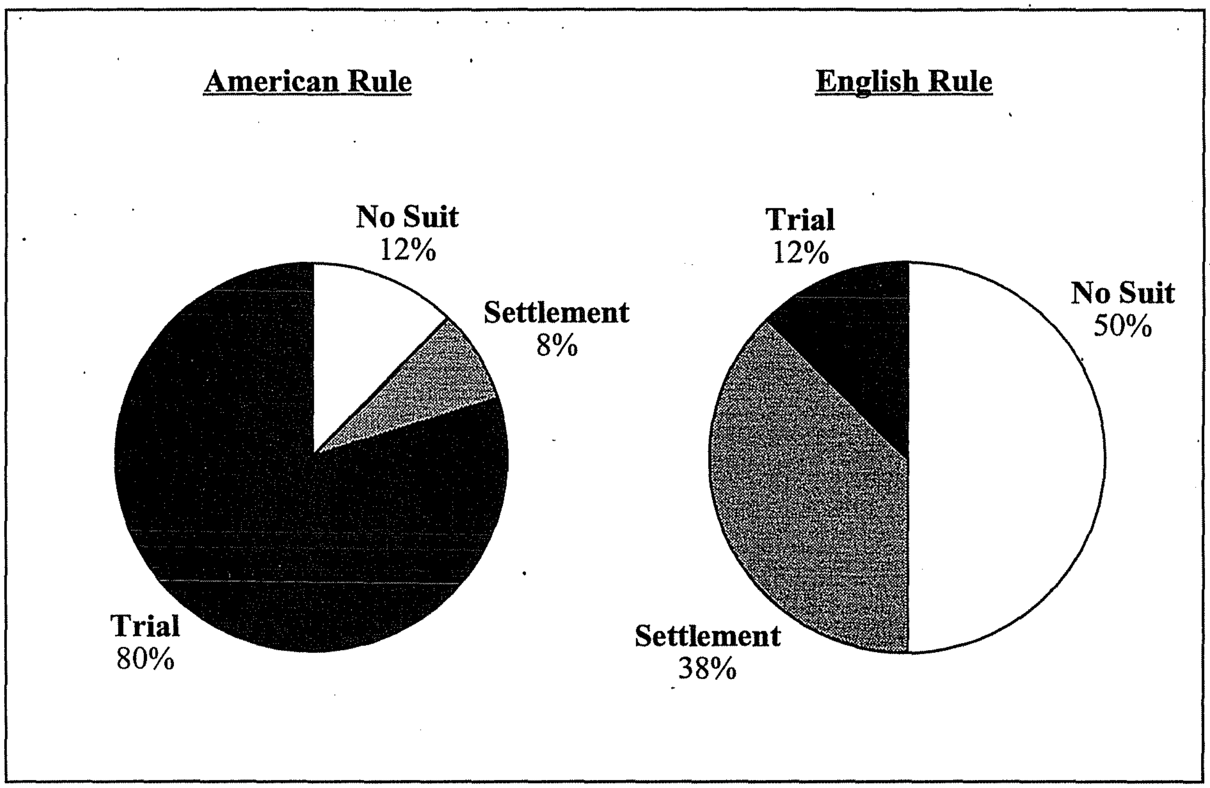


**Figure 11**

**Expenditure at Trial as a Function of Lawyer Productivity**

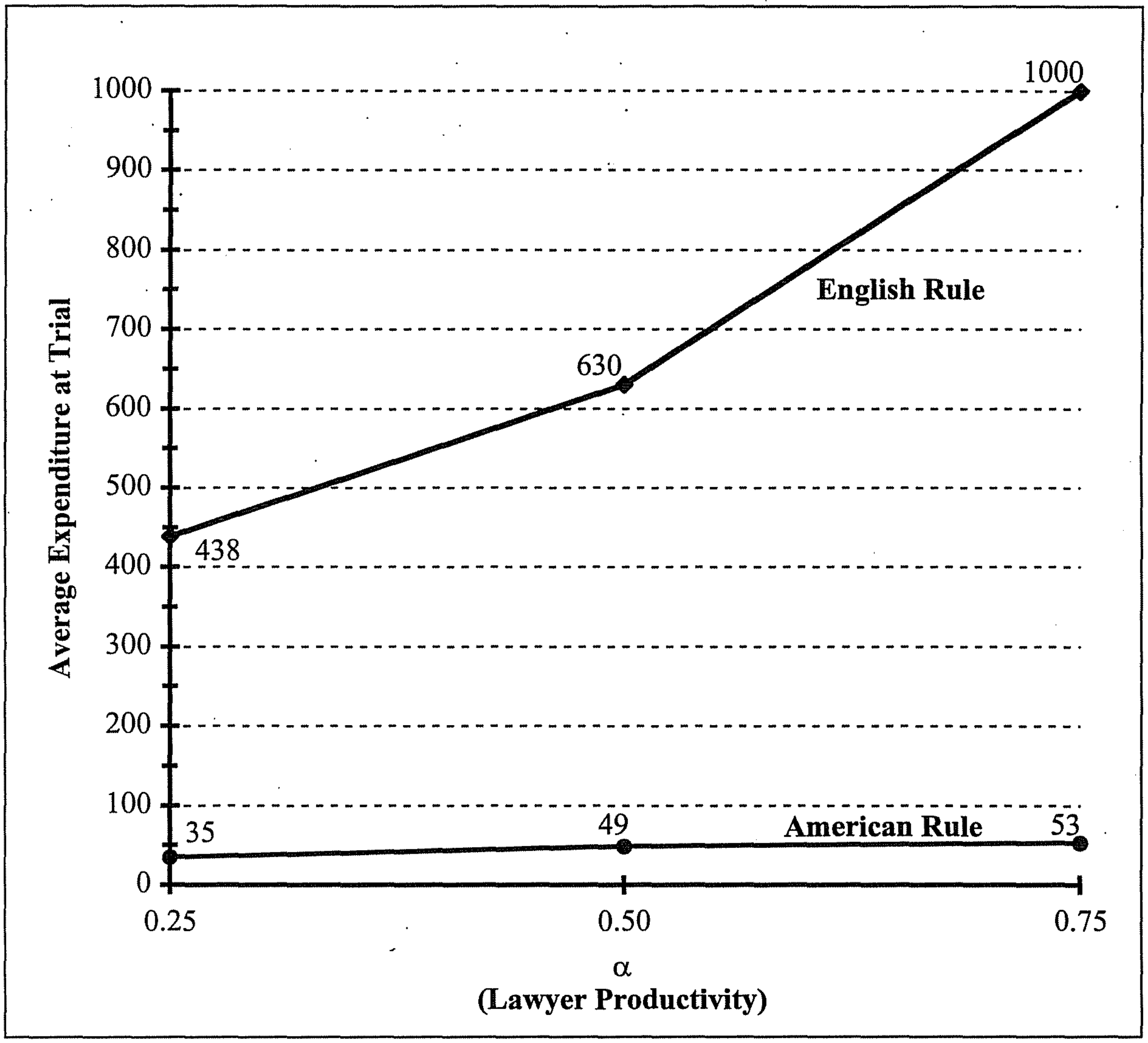


**Figure 12**

**Form of Resolution as a Function of Lawyer Productivity**

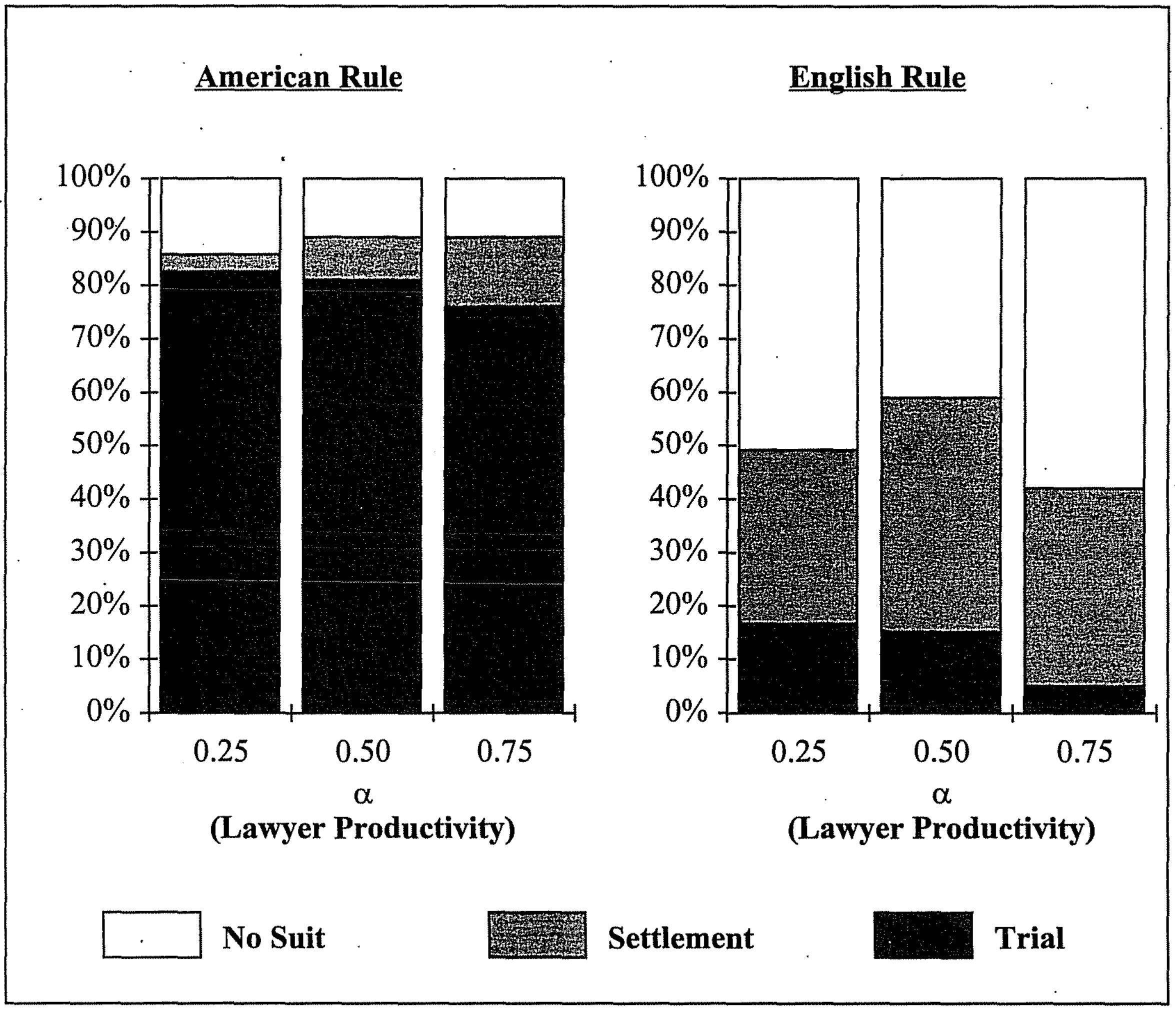


**Figure 13**

**Expenditure at Trial as a Function of Case Merit**

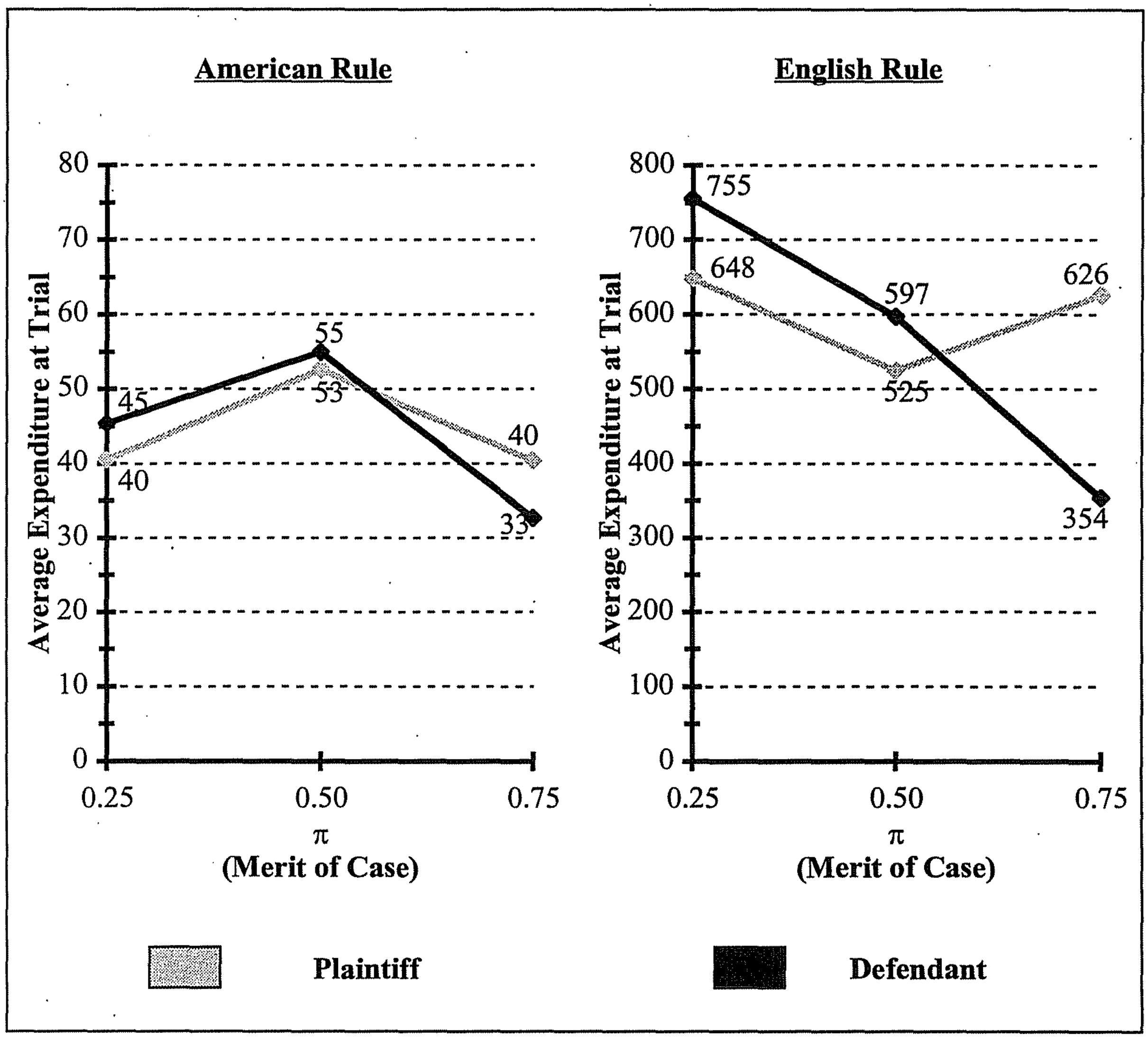


**Figure 14**

**Form of Resolution as a Function of Case Merit**

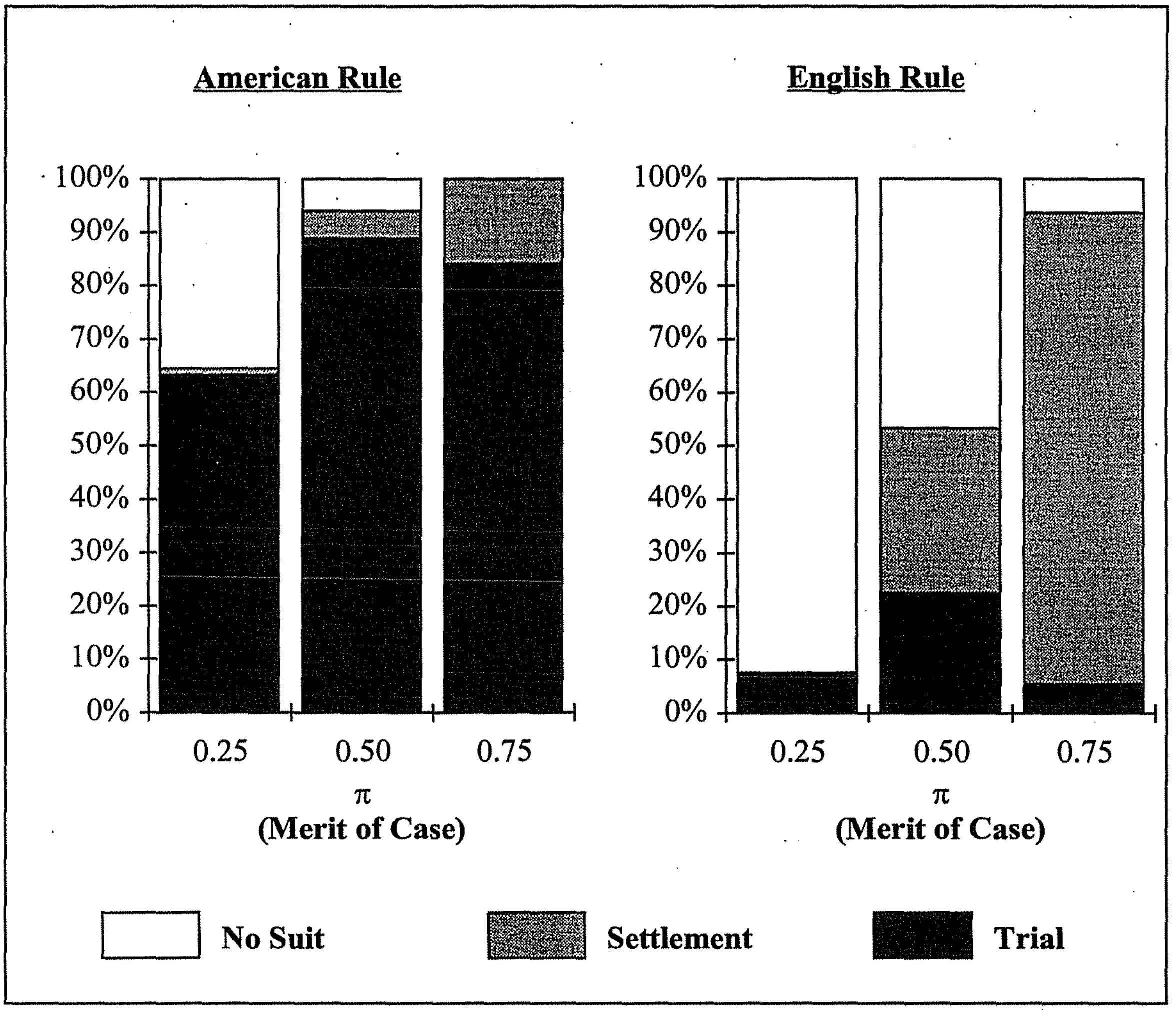


**Figure 15**

**Known Merit vs. Uncertain Merit Under the American Rule**

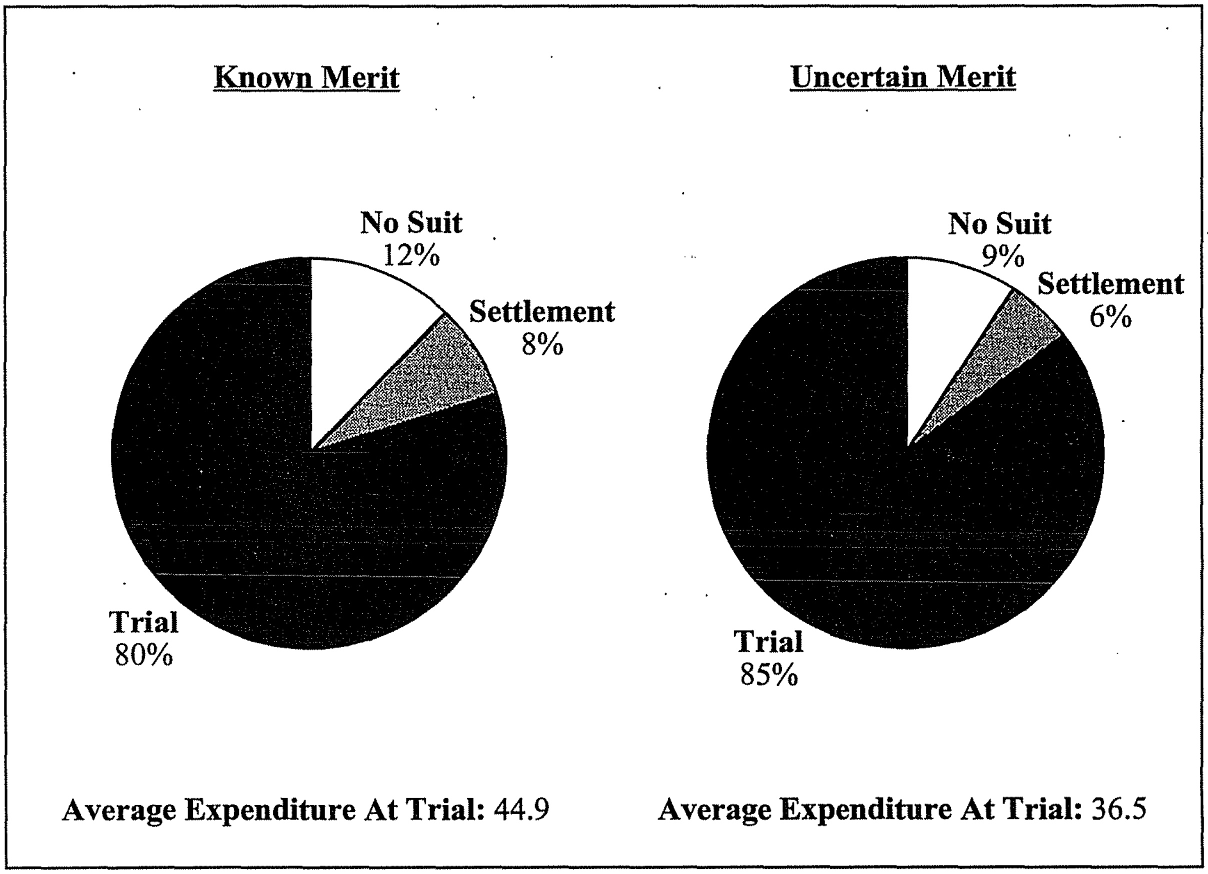


**Figure 16**

**Known Merit vs. Uncertain Merit Under the English Rule**

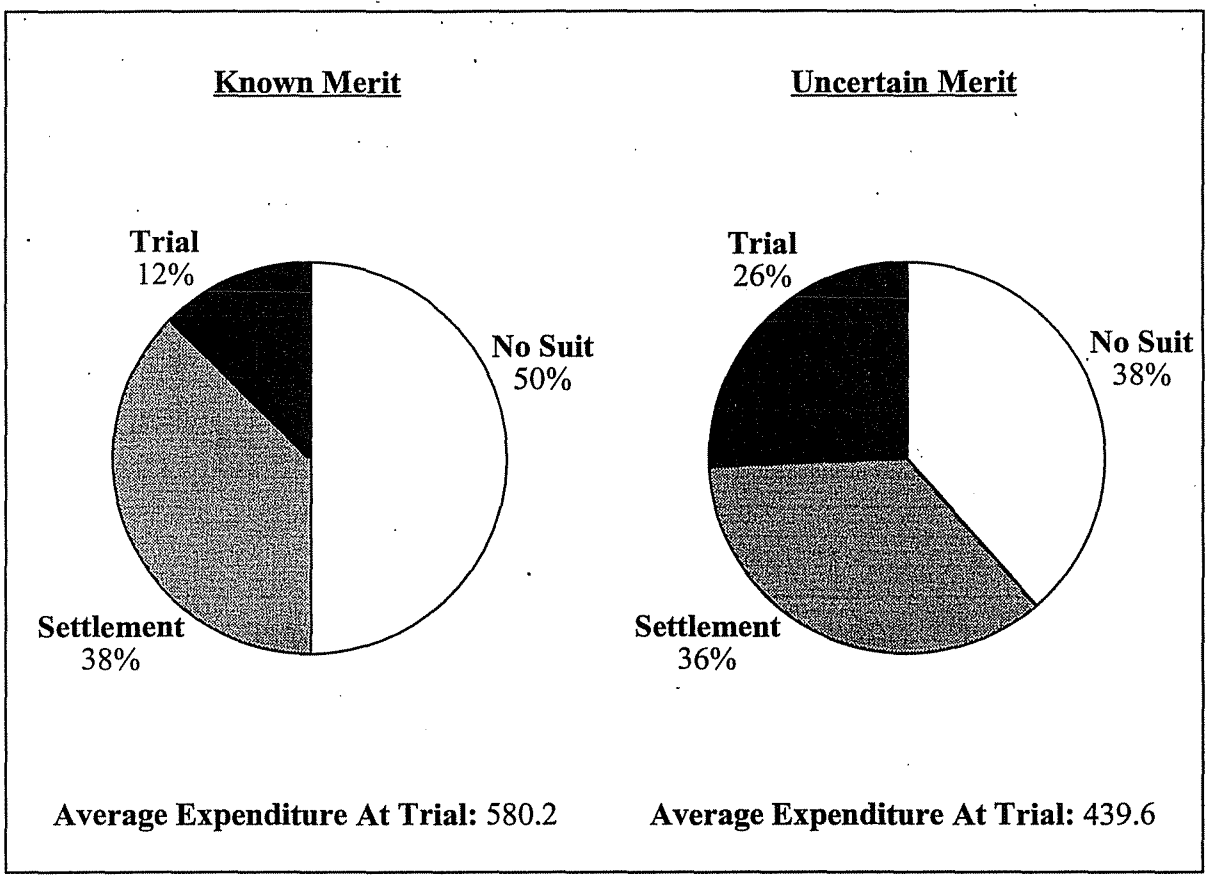


**Figure 17**

**Predicted vs. Observed Expenditure at Trial**

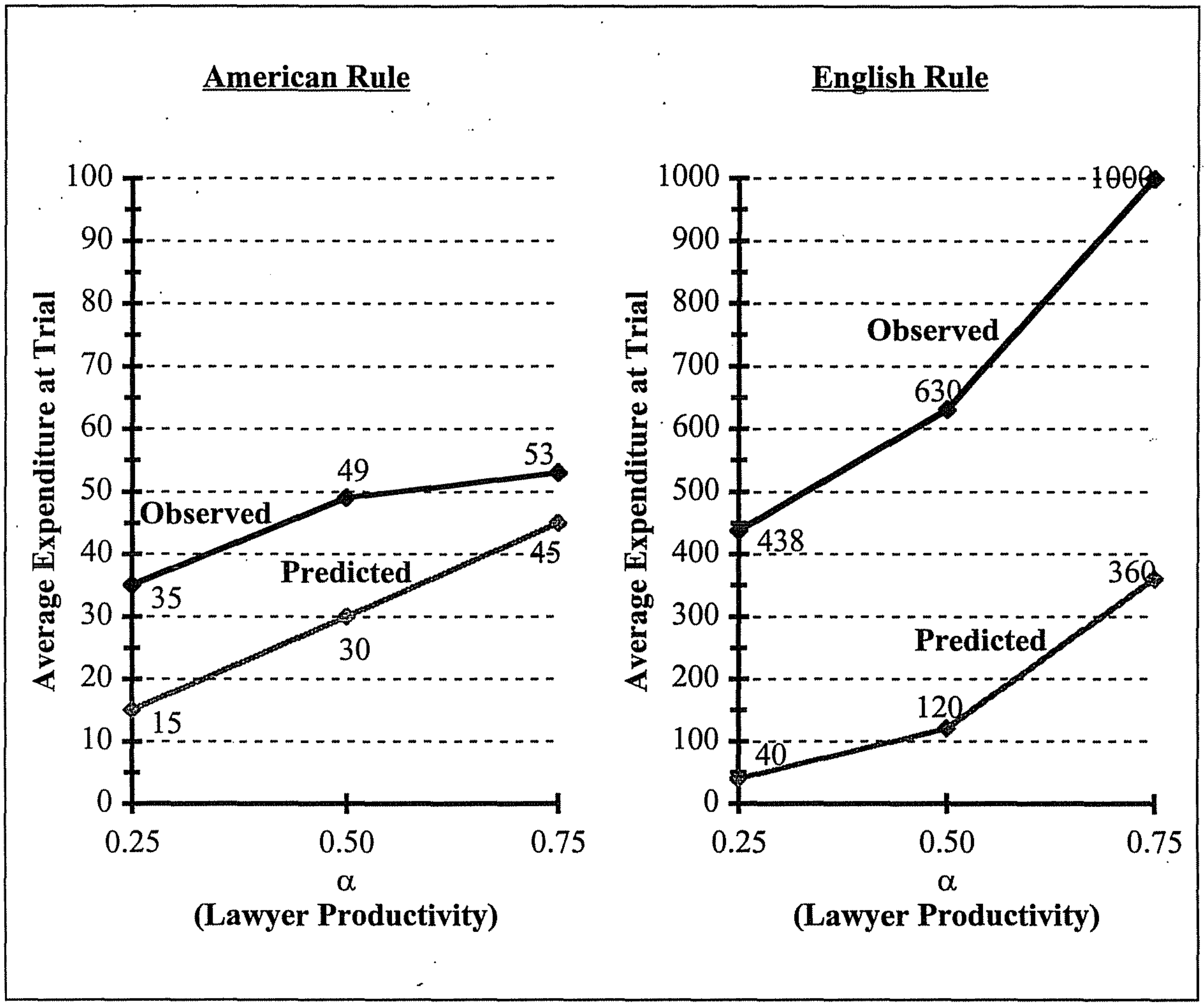


**Figure 18**

**Predicted vs. Observed Form of Resolution Under American Rule**

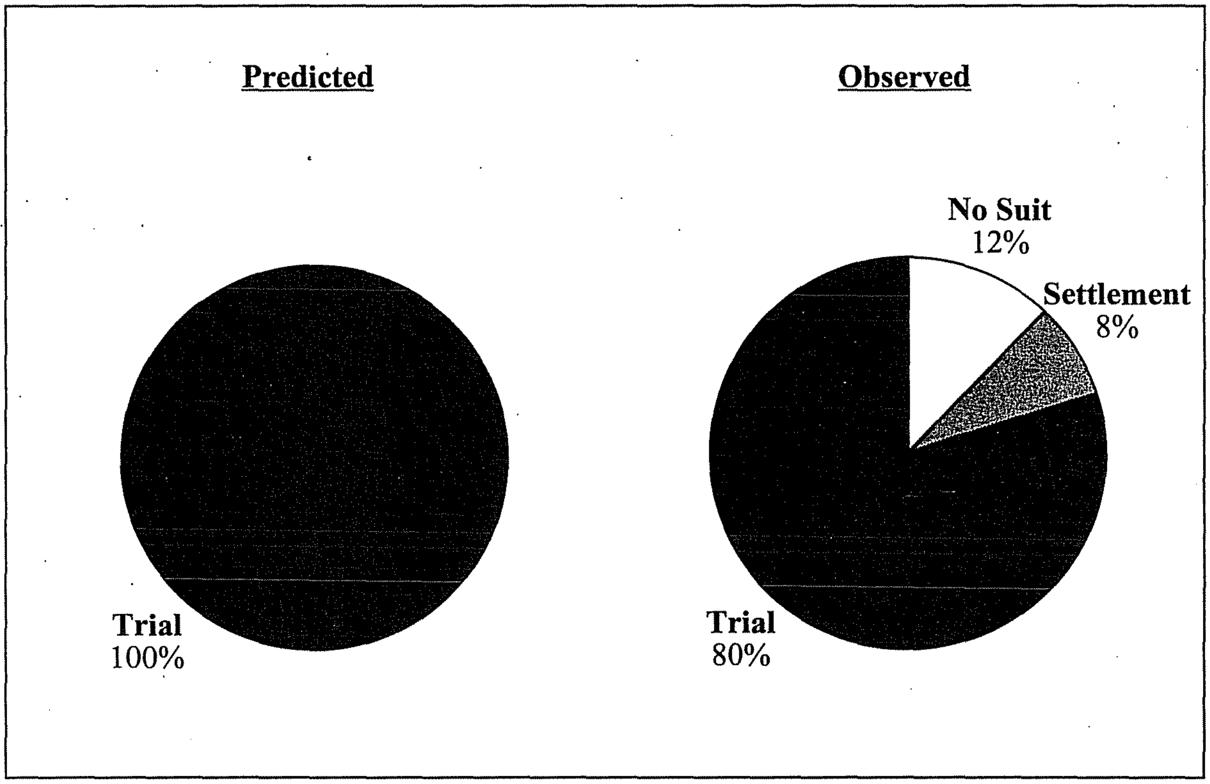
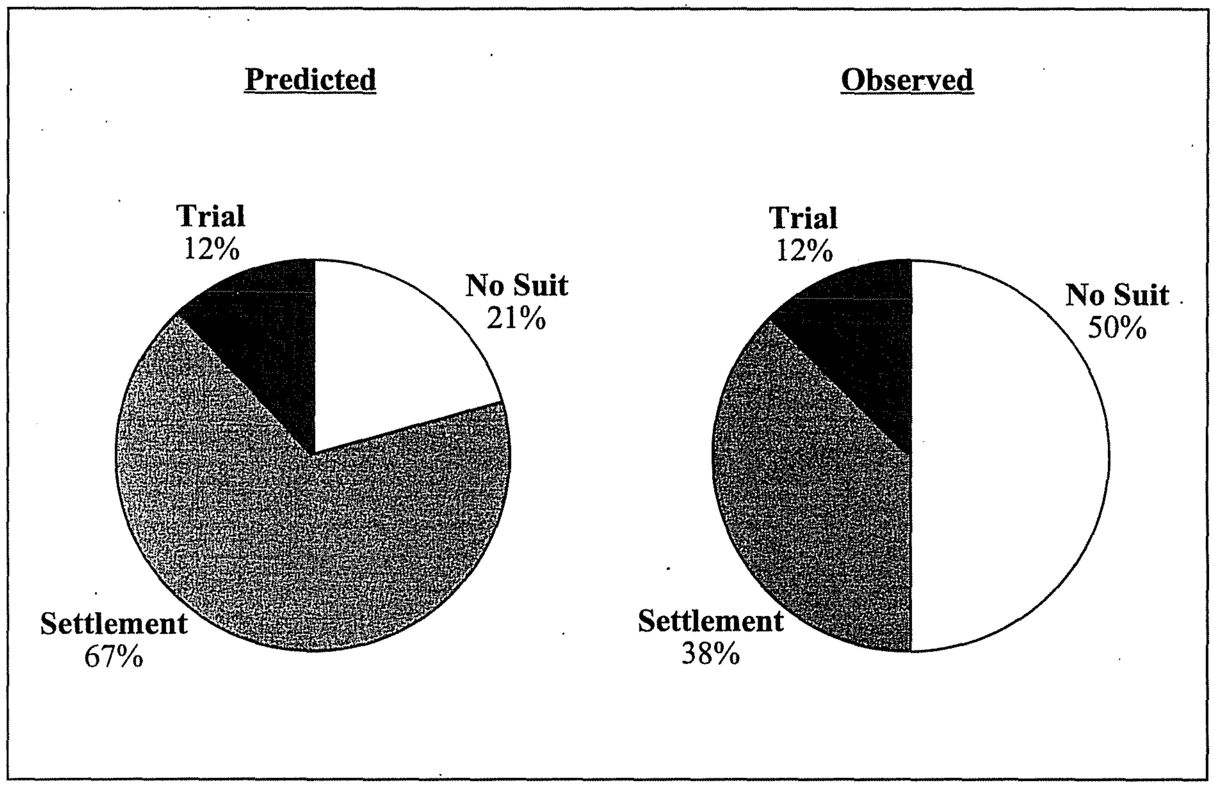


**Figure 19**

**Predicted vs. Observed Form of Resolution Under English Rule**



## APPENDIX: EXPERIMENT INSTRUCTIONS

This is an experiment in market decision making. If you follow the instructions carefully and make good decisions, you may earn money which will be paid to you in cash.

The currency used in this experiment is francs. Each franc is worth ____ dollars to you.

The experiment will consist of several periods. At the beginning of each period, every participant in the experiment will be randomly paired with another participant. In each period, you are equally likely to be paired with any other participant and the identity of the person you are paired with will never be revealed to you.

One member of each pair will randomly be designated as *Person A*, and the other member will be designated as *Person B*. In addition, each pair will randomly be assigned a *State*. The three possible states are "X," "Y," and "Z," and they each occur with equal probability. You will not know which State your pair has been assigned until the end of the period.

After each pair has been assigned a State, Person A and Person B will each receive a *Signal*. The Signal Person A receives is known as *Signal A* and the Signal Person B receives is known as *Signal B*. The three possible Signals are "X," "Y," and "Z." The probability of receiving each Signal will depend on which State the pair has been assigned. The following chart identifies the probability of receiving each Signal, "X," "Y," or "Z," as a function of the State, "X," "Y," or "Z," which has been assigned to the pair:

| **State** | **Signal** | | |
|---|---|---|---|
| | **X** | **Y** | **Z** |
| **X** | 60% | 20% | 20% |
| **Y** | 20% | 60% | 20% |
| **Z** | 20% | 20% | 60% |

In other words, each person has a 60% chance of receiving the Signal which matches the State the pair has been assigned, and a 20% chance of receiving each of the other two signals. For every pair, Person A and Person B will each be assigned a Signal according to the above probabilities. Thus, Person A and Person B could receive the same Signal or they could receive different Signals.

At the start of each period, the first thing you will see on the computer screen will be an identification of which Person you are and which Signal you have received. For example, if you are Person A and you have received Signal X, the computer screen will read: "You are Person A in group, your Signal is X."

Each participant will receive a *Capital Payment* of 400 francs at the beginning of each period. During the rest of the period, participants will make decisions that affect their *Period Payoff*. Each participant's final *Period Profit or Loss* will be the 400 franc Capital Payment plus or minus this Period Payoff.

Each period will consist of two stages:

**Stage 1**

At the beginning of Stage 1, Person A in each group will be asked: "Do you want to continue (Y/N)?" Person A can answer this question by pressing either "Y" or "N" on his or her keyboard.

If Person A chooses "N," the period ends for that pair. Both Person A and Person B will receive a Period Payoff of 0 francs. Therefore, they each will have a Period Profit of 400 francs (the Capital Payment of 400 francs plus the Period Payoff of 0 francs).

If Person A chooses "Y," he or she will pay a *Fee* of 10 francs for choosing to continue, and Person B is then asked the same question, "Do you want to continue (Y/N)?"

If Person B then chooses "N," Person B gives Person A a *Transfer* of 240 francs and the period ends for that pair. Thus, Person A will receive a Period Payoff of 230 francs (the Transfer of 240 francs minus the Fee of 10 francs) and Person B will receive a Period Payoff of -240 francs. The Period Profits for this pair will be 630 francs and 160 francs respectively.

If Person B chooses "Y," he or she will also pay a fee of 10 francs for choosing to continue, and the period proceeds to Stage 2.

**Stage 2**

During Stage 2, Person A and Person B will make *Investment* decisions which will affect the likelihood of two possible outcomes: *Outcome A* and *Outcome B.* Under Outcome A, Person B will give Person A a Transfer of 240 francs. Under Outcome B, no transfer takes place.

At the beginning of stage 2, each person is asked "Please enter your level of investment followed by the [F1] key to send." At this point each person will enter the amount of francs he or she wants to invest to affect the likelihood of Outcome A and Outcome B. The amount Person A invests is known as *Investment A* and the amount Person B invests is known as *Investment B.* Each person may enter any amount between 0 and 1000 (Note: You may invest more than your Capital Payment of 400 francs and you may also invest as little as 0 francs). After the amount is entered, you must press the F1 key to tell the computer you are ready.

The exact manner in which Investment A and Investment B affect the likelihood of Outcome A and Outcome B will be discussed in the final section of the instructions.

**Calculating Profits and/or Losses**

After stage 1 and stage 2 have been completed, and outcomes are calculated for each pair, every participant will be notified of the final results for his or her pair. For example, if you are Person B and you received Signal Z, at the end of the period your computer screen might read:

| | | |
|---|---|---|
| Period Ended. | State: X | Outcome: A |
| | You: B | Other: A |
| Signal: | Z | X |
| Invest: | 60 | 30 |
| Payoff: | -310 | 200 |

In the above case, the State was X, Person A received Signal X, and Person B received Signal Z. Both Person A and Person B chose to continue in Stage 1, Person B chose to invest 60 in Stage 2, and Person A chose to invest 30 in Stage 2. Since the outcome was Outcome A, Person B's Period Payoff is -310 (the Transfer of 240, the Investment of 60, and the Fee for continuing of 10) and Person A's payoff is 200 (the Transfer of 240 minus the Investment of 30 and the Fee of 10).

Period payoffs can be summarized by the following table:

| Stage 1 Decisions | | Outcome | Period Payoffs | |
|---|---|---|---|---|
| Person A | Person B | | Person A | Person B |
| N | - | - | 0 | 0 |
| Y | N | - | 230 | -240 |
| Y | Y | A | 230 - Investment A | -250 - Investment B |
| Y | Y | B | -10 - Investment A | -10 - Investment B |

At the end of each period, participants should fill out all of the columns of information on the Profit / Loss Record sheet and calculate their Period Profit or Loss by adding their Period Payoff to their Capital Payment of 400 francs.

**Determining the Outcome of Stage 2 for Each Pair**

The outcome of Stage 2 for each pair will be determined by a single draw from a computerized urn. The exact make-up of the urn will be determined by the investment decisions of the two individuals.

The urn is filled with 1000 balls. The first 500 balls will be divided proportionately between Person A and Person B based on the amount of francs each person has chosen to invest. In other words:

$$\text{Number of Balls assigned to Person A} = \frac{\text{Investment A}}{\text{Investment A} + \text{Investment B}} \times 500$$

$$\text{Number of Balls assigned to Person B} = \frac{\text{Investment B}}{\text{Investment A} + \text{Investment B}} \times 500$$

To better understand this, here are a few examples:

| Investment A | Investment B | Balls Assigned to Person A | Balls Assigned to Person B |
|---|---|---|---|
| 75 | 75 | 250 | 250 |
| 20 | 30 | 200 | 300 |
| 120 | 30 | 400 | 100 |
| 0 | 5 | 0 | 500 |
| 0 | 0 | 250 | 250 |

The assignment of the remaining 500 balls will be determined by the state, "X," "Y," or "Z." The following chart summarizes the assignment of these 500 balls:

| State | Balls Assigned to Person A | Balls Assigned to Person B |
|---|---|---|
| X | 125 | 375 |
| Y | 250 | 250 |
| Z | 375 | 125 |

After all 1000 balls have been assigned, a single ball is drawn from the urn. If the ball belongs to Person A, then Outcome A occurs and Person B transfers 240 francs to Person A. If the ball belongs to Person B, then Outcome B occurs and no transfer takes place.

To help you better understand how the 1000 balls are assigned, you have been provided three sheets labeled "Probability of Outcome A as a Function of Investment A and Investment B." Each sheet is also labelled either "State X," "State Y," or "State Z." These sheets each contain a chart which indicates the probability of Outcome A (or percentage of balls assigned to Person A) for combinations of Investment A and Investment B.

After examining the charts on these three sheets, please note the following observations:

(1) For a given amount of investment by Person B, the more Person A invests, the more likely Outcome A is and the less likely Outcome B is. Similiarly, for a given amount of investment by Person A, the more Person B invests, the more likely Outcome B is and the less likely Outcome A is.

(2) For any given combination of Investment A and Investment B, Outcome A is most likely in State Z and least likely in State X.

(3) In State X, no matter how much Person A invests, there is always at least a 37.5% chance of Outcome B. No matter how much Person B invests, there is always at least a 12.5% chance of Outcome A. Similarly, in State Y, there is always at least a 25% chance of Outcome B and there is always at least a 25% chance of Outcome A. In State Z, there is always at least a 12.5% chance of Outcome B and there is always at least a 37.5% chance of Outcome A.

Probability of Outcome A as a Function of Investment A and Investment B

*S T A T E X*

**I N V E S T M E N T B**

| INVESTMENT A | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 | 110 | 120 | 130 | 140 | 150 | 160 | 170 | 180 | 190 | 200 | 210 | 220 | 230 | 240 | 250 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 38 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 |
| 10 | 63 | 38 | 29 | 25 | 23 | 21 | 20 | 19 | 18 | 18 | 17 | 17 | 16 | 16 | 16 | 16 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 14 |
| 20 | 63 | 46 | 38 | 33 | 29 | 27 | 25 | 24 | 23 | 22 | 21 | 20 | 20 | 19 | 19 | 18 | 18 | 18 | 18 | 17 | 17 | 17 | 17 | 17 | 16 | 16 |
| 30 | 63 | 50 | 43 | 38 | 34 | 31 | 29 | 28 | 26 | 25 | 24 | 23 | 23 | 22 | 21 | 21 | 20 | 20 | 20 | 19 | 19 | 19 | 19 | 18 | 18 | 18 |
| 40 | 63 | 53 | 46 | 41 | 38 | 35 | 33 | 31 | 29 | 28 | 27 | 26 | 25 | 24 | 24 | 23 | 23 | 22 | 22 | 21 | 21 | 21 | 20 | 20 | 20 | 19 |
| 50 | 63 | 54 | 48 | 44 | 40 | 38 | 35 | 33 | 32 | 30 | 29 | 28 | 27 | 26 | 26 | 25 | 24 | 24 | 23 | 23 | 23 | 22 | 22 | 21 | 21 | 21 |
| 60 | 63 | 55 | 50 | 46 | 43 | 40 | 38 | 36 | 34 | 33 | 31 | 30 | 29 | 28 | 28 | 27 | 26 | 26 | 25 | 25 | 24 | 24 | 23 | 23 | 23 | 22 |
| 70 | 63 | 56 | 51 | 48 | 44 | 42 | 39 | 38 | 36 | 34 | 33 | 32 | 31 | 30 | 29 | 28 | 28 | 27 | 27 | 26 | 25 | 25 | 25 | 24 | 24 | 23 |
| 80 | 63 | 57 | 53 | 49 | 46 | 43 | 41 | 39 | 38 | 36 | 35 | 34 | 33 | 32 | 31 | 30 | 29 | 29 | 28 | 27 | 27 | 26 | 26 | 25 | 25 | 25 |
| 90 | 63 | 58 | 53 | 50 | 47 | 45 | 43 | 41 | 39 | 38 | 36 | 35 | 34 | 33 | 32 | 31 | 31 | 30 | 29 | 29 | 28 | 28 | 27 | 27 | 26 | 26 |
| 100 | 63 | 58 | 54 | 51 | 48 | 46 | 44 | 42 | 40 | 39 | 38 | 36 | 35 | 34 | 33 | 33 | 32 | 31 | 30 | 30 | 29 | 29 | 28 | 28 | 27 | 27 |
| 110 | 63 | 58 | 55 | 52 | 49 | 47 | 45 | 43 | 41 | 40 | 39 | 38 | 36 | 35 | 35 | 34 | 33 | 32 | 31 | 31 | 30 | 30 | 29 | 29 | 28 | 28 |
| 120 | 63 | 59 | 55 | 53 | 50 | 48 | 46 | 44 | 43 | 41 | 40 | 39 | 38 | 37 | 36 | 35 | 34 | 33 | 33 | 32 | 31 | 31 | 30 | 30 | 29 | 29 |
| 130 | 63 | 59 | 56 | 53 | 51 | 49 | 47 | 45 | 43 | 42 | 41 | 40 | 39 | 38 | 37 | 36 | 35 | 34 | 33 | 33 | 32 | 32 | 31 | 31 | 30 | 30 |
| 140 | 63 | 59 | 56 | 54 | 51 | 49 | 48 | 46 | 44 | 43 | 42 | 41 | 39 | 38 | 38 | 37 | 36 | 35 | 34 | 34 | 33 | 33 | 32 | 31 | 31 | 30 |
| 150 | 63 | 59 | 57 | 54 | 52 | 50 | 48 | 47 | 45 | 44 | 43 | 41 | 40 | 39 | 38 | 38 | 37 | 36 | 35 | 35 | 34 | 33 | 33 | 32 | 32 | 31 |
| 160 | 63 | 60 | 57 | 55 | 53 | 51 | 49 | 47 | 46 | 45 | 43 | 42 | 41 | 40 | 39 | 38 | 38 | 37 | 36 | 35 | 35 | 34 | 34 | 33 | 33 | 32 |
| 170 | 63 | 60 | 57 | 55 | 53 | 51 | 49 | 48 | 47 | 45 | 44 | 43 | 42 | 41 | 40 | 39 | 38 | 38 | 37 | 36 | 35 | 35 | 34 | 34 | 33 | 33 |
| 180 | 63 | 60 | 58 | 55 | 53 | 52 | 50 | 49 | 47 | 46 | 45 | 44 | 43 | 42 | 41 | 40 | 39 | 38 | 38 | 37 | 36 | 36 | 35 | 34 | 34 | 33 |
| 190 | 63 | 60 | 58 | 56 | 54 | 52 | 51 | 49 | 48 | 46 | 45 | 44 | 43 | 42 | 41 | 40 | 40 | 39 | 38 | 38 | 37 | 36 | 36 | 35 | 35 | 34 |
| 200 | 63 | 60 | 58 | 56 | 54 | 53 | 51 | 50 | 48 | 47 | 46 | 45 | 44 | 43 | 42 | 41 | 40 | 40 | 39 | 38 | 38 | 37 | 36 | 36 | 35 | 35 |
| 210 | 63 | 60 | 58 | 56 | 55 | 53 | 51 | 50 | 49 | 48 | 46 | 45 | 44 | 43 | 43 | 42 | 41 | 40 | 39 | 39 | 38 | 38 | 37 | 36 | 36 | 35 |
| 220 | 63 | 60 | 58 | 57 | 55 | 53 | 52 | 50 | 49 | 48 | 47 | 46 | 45 | 44 | 43 | 42 | 41 | 41 | 40 | 39 | 39 | 38 | 38 | 37 | 36 | 36 |
| 230 | 63 | 60 | 59 | 57 | 55 | 54 | 52 | 51 | 50 | 48 | 47 | 46 | 45 | 44 | 44 | 43 | 42 | 41 | 41 | 40 | 39 | 39 | 38 | 38 | 37 | 36 |
| 240 | 63 | 61 | 59 | 57 | 55 | 54 | 53 | 51 | 50 | 49 | 48 | 47 | 46 | 45 | 44 | 43 | 43 | 42 | 41 | 40 | 40 | 39 | 39 | 38 | 38 | 37 |
| 250 | 63 | 61 | 59 | 57 | 56 | 54 | 53 | 52 | 50 | 49 | 48 | 47 | 46 | 45 | 45 | 44 | 43 | 42 | 42 | 41 | 40 | 40 | 39 | 39 | 38 | 38 |

Prob(A) > 60

40 ≤ Prob(A) ≤ 60

Prob(A) < 40

Probability of Outcome A as a Function of Investment A and Investment B

*S T A T E   Y*

**I N V E S T M E N T   B**

**I N V E S T M E N T   A** (rows)

| | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 | 110 | 120 | 130 | 140 | 150 | 160 | 170 | 180 | 190 | 200 | 210 | 220 | 230 | 240 | 250 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 50 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 |
| 10 | 75 | 50 | 42 | 38 | 35 | 33 | 32 | 31 | 31 | 30 | 30 | 29 | 29 | 29 | 28 | 28 | 28 | 28 | 28 | 28 | 27 | 27 | 27 | 27 | 27 | 27 |
| 20 | 75 | 58 | 50 | 45 | 42 | 39 | 38 | 36 | 35 | 34 | 33 | 33 | 32 | 32 | 31 | 31 | 31 | 30 | 30 | 30 | 30 | 29 | 29 | 29 | 29 | 29 |
| 30 | 75 | 63 | 55 | 50 | 46 | 44 | 42 | 40 | 39 | 38 | 37 | 36 | 35 | 34 | 34 | 33 | 33 | 33 | 32 | 32 | 32 | 31 | 31 | 31 | 31 | 30 |
| 40 | 75 | 65 | 58 | 54 | 50 | 47 | 45 | 43 | 42 | 40 | 39 | 38 | 38 | 37 | 36 | 36 | 35 | 35 | 34 | 34 | 33 | 33 | 33 | 32 | 32 | 32 |
| 50 | 75 | 67 | 61 | 56 | 53 | 50 | 48 | 46 | 44 | 43 | 42 | 41 | 40 | 39 | 38 | 38 | 37 | 36 | 36 | 35 | 35 | 35 | 34 | 34 | 34 | 33 |
| 60 | 75 | 68 | 63 | 58 | 55 | 52 | 50 | 48 | 46 | 45 | 44 | 43 | 42 | 41 | 40 | 39 | 39 | 38 | 38 | 37 | 37 | 36 | 36 | 35 | 35 | 35 |
| 70 | 75 | 69 | 64 | 60 | 57 | 54 | 52 | 50 | 48 | 47 | 46 | 44 | 43 | 43 | 42 | 41 | 40 | 40 | 39 | 38 | 38 | 38 | 37 | 37 | 36 | 36 |
| 80 | 75 | 69 | 65 | 61 | 58 | 56 | 54 | 52 | 50 | 49 | 47 | 46 | 45 | 44 | 43 | 42 | 42 | 41 | 40 | 40 | 39 | 39 | 38 | 38 | 38 | 37 |
| 90 | 75 | 70 | 66 | 63 | 60 | 57 | 55 | 53 | 51 | 50 | 49 | 48 | 46 | 45 | 45 | 44 | 43 | 42 | 42 | 41 | 41 | 40 | 40 | 39 | 39 | 38 |
| 100 | 75 | 70 | 67 | 63 | 61 | 58 | 56 | 54 | 53 | 51 | 50 | 49 | 48 | 47 | 46 | 45 | 44 | 44 | 43 | 42 | 42 | 41 | 41 | 40 | 40 | 39 |
| 110 | 75 | 71 | 67 | 64 | 62 | 59 | 57 | 56 | 54 | 53 | 51 | 50 | 49 | 48 | 47 | 46 | 45 | 45 | 44 | 43 | 43 | 42 | 42 | 41 | 41 | 40 |
| 120 | 75 | 71 | 68 | 65 | 63 | 60 | 58 | 57 | 55 | 54 | 52 | 51 | 50 | 49 | 48 | 47 | 46 | 46 | 45 | 44 | 44 | 43 | 43 | 42 | 42 | 41 |
| 130 | 75 | 71 | 68 | 66 | 63 | 61 | 59 | 58 | 56 | 55 | 53 | 52 | 51 | 50 | 49 | 48 | 47 | 47 | 46 | 45 | 45 | 44 | 44 | 43 | 43 | 42 |
| 140 | 75 | 72 | 69 | 66 | 64 | 62 | 60 | 58 | 57 | 55 | 54 | 53 | 52 | 51 | 50 | 49 | 48 | 48 | 47 | 46 | 46 | 45 | 44 | 44 | 43 | 43 |
| 150 | 75 | 72 | 69 | 67 | 64 | 63 | 61 | 59 | 58 | 56 | 55 | 54 | 53 | 52 | 51 | 50 | 49 | 48 | 48 | 47 | 46 | 46 | 45 | 45 | 44 | 44 |
| 160 | 75 | 72 | 69 | 67 | 65 | 63 | 61 | 60 | 58 | 57 | 56 | 55 | 54 | 53 | 52 | 51 | 50 | 49 | 49 | 48 | 47 | 47 | 46 | 46 | 45 | 45 |
| 170 | 75 | 72 | 70 | 68 | 65 | 64 | 62 | 60 | 59 | 58 | 56 | 55 | 54 | 53 | 52 | 52 | 51 | 50 | 49 | 49 | 48 | 47 | 47 | 46 | 46 | 45 |
| 180 | 75 | 72 | 70 | 68 | 66 | 64 | 63 | 61 | 60 | 58 | 57 | 56 | 55 | 54 | 53 | 52 | 51 | 51 | 50 | 49 | 49 | 48 | 48 | 47 | 46 | 46 |
| 190 | 75 | 73 | 70 | 68 | 66 | 65 | 63 | 62 | 60 | 59 | 58 | 57 | 56 | 55 | 54 | 53 | 52 | 51 | 51 | 50 | 49 | 49 | 48 | 48 | 47 | 47 |
| 200 | 75 | 73 | 70 | 68 | 67 | 65 | 63 | 62 | 61 | 59 | 58 | 57 | 56 | 55 | 54 | 54 | 53 | 52 | 51 | 51 | 50 | 49 | 49 | 48 | 48 | 47 |
| 210 | 75 | 73 | 71 | 69 | 67 | 65 | 64 | 63 | 61 | 60 | 59 | 58 | 57 | 56 | 55 | 54 | 53 | 53 | 52 | 51 | 51 | 50 | 49 | 49 | 48 | 48 |
| 220 | 75 | 73 | 71 | 69 | 67 | 66 | 64 | 63 | 62 | 60 | 59 | 58 | 57 | 56 | 56 | 55 | 54 | 53 | 53 | 52 | 51 | 51 | 50 | 49 | 49 | 48 |
| 230 | 75 | 73 | 71 | 69 | 68 | 66 | 65 | 63 | 62 | 61 | 60 | 59 | 58 | 57 | 56 | 55 | 54 | 54 | 53 | 52 | 52 | 51 | 51 | 50 | 49 | 49 |
| 240 | 75 | 73 | 71 | 69 | 68 | 66 | 65 | 64 | 63 | 61 | 60 | 59 | 58 | 57 | 57 | 56 | 55 | 54 | 54 | 53 | 52 | 52 | 51 | 51 | 50 | 49 |
| 250 | 75 | 73 | 71 | 70 | 68 | 67 | 65 | 64 | 63 | 62 | 61 | 60 | 59 | 58 | 57 | 56 | 55 | 55 | 54 | 53 | 53 | 52 | 52 | 51 | 51 | 50 |

Prob(A) > 60 — 40 ≤ Prob(A) ≤ 60 — Prob(A) < 40

Probability of Outcome A as a Function of Investment A and Investment B

*S T A T E  Z*

**I N V E S T M E N T  B**

| | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 | 110 | 120 | 130 | 140 | 150 | 160 | 170 | 180 | 190 | 200 | 210 | 220 | 230 | 240 | 250 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 63 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 38 |
| 10 | 88 | 63 | 54 | 50 | 48 | 46 | 45 | 44 | 43 | 43 | 42 | 42 | 41 | 41 | 41 | 41 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 40 | 39 |
| 20 | 88 | 71 | 63 | 58 | 54 | 52 | 50 | 49 | 48 | 47 | 46 | 45 | 45 | 44 | 44 | 43 | 43 | 43 | 43 | 42 | 42 | 42 | 42 | 42 | 41 | 41 |
| 30 | 88 | 75 | 68 | 63 | 59 | 56 | 54 | 53 | 51 | 50 | 49 | 48 | 48 | 47 | 46 | 46 | 45 | 45 | 45 | 44 | 44 | 44 | 44 | 43 | 43 | 43 |
| 40 | 88 | 78 | 71 | 66 | 63 | 60 | 58 | 56 | 54 | 53 | 52 | 51 | 50 | 49 | 49 | 48 | 48 | 47 | 47 | 46 | 46 | 46 | 45 | 45 | 45 | 44 |
| 50 | 88 | 79 | 73 | 69 | 65 | 63 | 60 | 58 | 57 | 55 | 54 | 53 | 52 | 51 | 51 | 50 | 49 | 49 | 48 | 48 | 48 | 47 | 47 | 46 | 46 | 46 |
| 60 | 88 | 80 | 75 | 71 | 68 | 65 | 63 | 61 | 59 | 58 | 56 | 55 | 54 | 53 | 53 | 52 | 51 | 51 | 50 | 50 | 49 | 49 | 48 | 48 | 48 | 47 |
| 70 | 88 | 81 | 76 | 73 | 69 | 67 | 64 | 63 | 61 | 59 | 58 | 57 | 56 | 55 | 54 | 53 | 53 | 52 | 52 | 51 | 50 | 50 | 50 | 49 | 49 | 48 |
| 80 | 88 | 82 | 78 | 74 | 71 | 68 | 66 | 64 | 63 | 61 | 60 | 59 | 58 | 57 | 56 | 55 | 54 | 54 | 53 | 52 | 52 | 51 | 51 | 50 | 50 | 50 |
| 90 | 88 | 83 | 78 | 75 | 72 | 70 | 68 | 66 | 64 | 63 | 61 | 60 | 59 | 58 | 57 | 56 | 56 | 55 | 54 | 54 | 53 | 53 | 52 | 52 | 51 | 51 |
| 100 | 88 | 83 | 79 | 76 | 73 | 71 | 69 | 67 | 65 | 64 | 63 | 61 | 60 | 59 | 58 | 58 | 57 | 56 | 55 | 55 | 54 | 54 | 53 | 53 | 52 | 52 |
| 110 | 88 | 83 | 80 | 77 | 74 | 72 | 70 | 68 | 66 | 65 | 64 | 63 | 61 | 60 | 60 | 59 | 58 | 57 | 56 | 56 | 55 | 55 | 54 | 54 | 53 | 53 |
| 120 | 88 | 84 | 80 | 78 | 75 | 73 | 71 | 69 | 68 | 66 | 65 | 64 | 63 | 62 | 61 | 60 | 59 | 58 | 58 | 57 | 56 | 56 | 55 | 55 | 54 | 54 |
| 130 | 88 | 84 | 81 | 78 | 76 | 74 | 72 | 70 | 68 | 67 | 66 | 65 | 64 | 63 | 62 | 61 | 60 | 59 | 58 | 58 | 57 | 57 | 56 | 56 | 55 | 55 |
| 140 | 88 | 84 | 81 | 79 | 76 | 74 | 73 | 71 | 69 | 68 | 67 | 66 | 64 | 63 | 63 | 62 | 61 | 60 | 59 | 59 | 58 | 58 | 57 | 56 | 56 | 55 |
| 150 | 88 | 84 | 82 | 79 | 77 | 75 | 73 | 72 | 70 | 69 | 68 | 66 | 65 | 64 | 63 | 63 | 62 | 61 | 60 | 60 | 59 | 58 | 58 | 57 | 57 | 56 |
| 160 | 88 | 85 | 82 | 80 | 78 | 76 | 74 | 72 | 71 | 70 | 68 | 67 | 66 | 65 | 64 | 63 | 63 | 62 | 61 | 60 | 60 | 59 | 59 | 58 | 58 | 57 |
| 170 | 88 | 85 | 82 | 80 | 78 | 76 | 74 | 73 | 72 | 70 | 69 | 68 | 67 | 66 | 65 | 64 | 63 | 63 | 62 | 61 | 60 | 60 | 59 | 59 | 58 | 58 |
| 180 | 88 | 85 | 83 | 80 | 78 | 77 | 75 | 74 | 72 | 71 | 70 | 69 | 68 | 67 | 66 | 65 | 64 | 63 | 63 | 62 | 61 | 61 | 60 | 59 | 59 | 58 |
| 190 | 88 | 85 | 83 | 81 | 79 | 77 | 76 | 74 | 73 | 71 | 70 | 69 | 68 | 67 | 66 | 65 | 65 | 64 | 63 | 63 | 62 | 61 | 61 | 60 | 60 | 59 |
| 200 | 88 | 85 | 83 | 81 | 79 | 78 | 76 | 75 | 73 | 72 | 71 | 70 | 69 | 68 | 67 | 66 | 65 | 65 | 64 | 63 | 63 | 62 | 61 | 61 | 60 | 60 |
| 210 | 88 | 85 | 83 | 81 | 80 | 78 | 76 | 75 | 74 | 73 | 71 | 70 | 69 | 68 | 68 | 67 | 66 | 65 | 64 | 64 | 63 | 63 | 62 | 61 | 61 | 60 |
| 220 | 88 | 85 | 83 | 82 | 80 | 78 | 77 | 75 | 74 | 73 | 72 | 71 | 70 | 69 | 68 | 67 | 66 | 66 | 65 | 64 | 64 | 63 | 63 | 62 | 61 | 61 |
| 230 | 88 | 85 | 84 | 82 | 80 | 79 | 77 | 76 | 75 | 73 | 72 | 71 | 70 | 69 | 69 | 68 | 67 | 66 | 66 | 65 | 64 | 64 | 63 | 63 | 62 | 61 |
| 240 | 88 | 86 | 84 | 82 | 80 | 79 | 78 | 76 | 75 | 74 | 73 | 72 | 71 | 70 | 69 | 68 | 68 | 67 | 66 | 65 | 65 | 64 | 64 | 63 | 63 | 62 |
| 250 | 88 | 86 | 84 | 82 | 81 | 79 | 78 | 77 | 75 | 74 | 73 | 72 | 71 | 70 | 70 | 69 | 68 | 67 | 67 | 66 | 65 | 65 | 64 | 64 | 63 | 63 |

Prob(A) > 60

40 ≤ Prob(A) ≤ 60

Prob(A) < 40

## REFERENCES


Anderson, D.A. 1994. "Improving Settlement Devices: Rule 68 and Beyond." *Journal of Legal Studies* 23:225-246.

Bebchuk, L.A. 1984. "Litigation and Settlement Under Imperfect Information." *RAND Journal of Economics* 15:404-415.

Beckner, C.F., and Avery Katz. 1995. "The Incentive Effects of Litigation Fee Shifting When Legal Standards Are Uncertain." *International Review of Law and Economics* 15:205-224.

Boggs, D. 1991. "Comment on Donohue." *Law and Contemporary Problems* 54:223-226.

Bowles, R. 1987. "Settlement Range and Cost Allocation Rules: A Comment on Avery Katz's 'Measuring the Demand for Litigation: Is the English Rule Really Cheaper?'" *Journal of Law, Economics, and Organization* 3:177-184.

Braeutigam, R., B. Owen, and J. Panzar. 1984. "An Economic Analysis of Alternative Fee Shifting Systems." *Law and Contemporary Problems* 47:173-185.

Cooter, R.D., and D.L. Rubinfeld. 1989. "Economic Analysis of Legal Disputes and Their Resolution." *Journal of Economic Literature* 27:1067-1097.

Coursey, D.L., and L.R. Stanley. 1988. "Pretrial Bargaining Behavior Within the Shadow of the Law: Theory and Experimental Evidence." *International Review of Law and Economics* 8:161-179.

Donohue, J.J., III. 1991. "The Effects of Fee Shifting on the Settlement Rate: Theoretical Observations on Costs, Conflicts, and Contingency Fees." *Law and Contemporary Problems* 54:194-226.

Donohue, J.J., III. 1991. "Opting for the British Rule, or if Posner and Shavell Can't Remember the Coase Theorem, Who Will?" *Harvard Law Review* 104:1093-1119.

*The Economist*. 1992. "Order in the Tort." July 18, 10-13.

Gravelle, H.S.E. 1993. "The Efficiency Implications of Cost-Shifting Rules." *International Review of Law and Economics* 13:3-18.

Hause, J.C. 1989. "Indemnity, Settlement, and Litigation, or I'll Be Suing You." *Journal of Legal Studies* 18:157-179.

Hay, B.L. 1995. "Effort, Information, Settlement, Trial." *Journal of Legal Studies* 24:29-62.

Hersch, P.L. 1990. "Indemnity, Settlement, and Litigation: Comment and Extension." *Journal of Legal Studies* 19:235-241.

Hughes, J.W., and E.A. Snyder. 1990. "The English Rule for Allocating Legal Costs: Evidence Confronts Theory." *Journal of Law, Economics, and Organization* 6:345-380.

Hughes, J.W., and E.A. Snyder. 1995. "The English Rule for Allocating Legal Costs: Evidence Confronts Theory." *Journal of Law and Economics* 38:225-250.

Hyde, H.J. 1995. "Should the House Pass H.R. 988, H.R. 1058, and H.R. 956, Legal Reform Legislation?" *Congressional Digest* 74(5):154-156.

Spier, K. 1994. "Pretrial Bargaining and the Design of Fee-Shifting Rules." *RAND Journal of Economics* 25:1-18.

Thomas, R.E. 1994. "The Selection Effect, Judicial Error and Litigation Expenditures." Unpublished Manuscript, University of Florida.

Thomas, R.E. 1994. "The Trial Selection Effect Without the 50 Percent Rule: Some Experimental Evidence." *Journal of Legal Studies* 24:209-228.