

Voting in Search of the Public Good: the Probabilistic Logic of Majority Judgments

James Hawthorne

Dept. of Philosophy, University of Oklahoma

hawthorne@ou.edu

running head: Probabilistic Logic of Majority Judgments

key words: Condorcet Jury Theorem, democracy, majority, public good, probabilistic logic, voting

Abstract. I argue for an epistemic conception of voting, a conception on which the purpose of the ballot is at least in some cases to identify which of several policy proposals will best promote the public good. To support this view I first briefly investigate several notions of the kind of public good that public policy should promote. Then I examine the probability logic of voting as embodied in two very robust versions of the *Condorcet Jury Theorem* and some related results. These theorems show that if the number of voters or legislators is sufficiently large and the average of their individual propensities to select the better of two policy proposals is a little above random chance, and if each person votes his or her own best judgment (rather than in alliance with a block or faction), then the majority is extremely likely to select the better alternative. Here ‘better alternative’ means that policy or law that will best promote the public good. I also explicate a Convincing Majorities Theorem, which shows the extent to which the majority vote should provide evidence that the better policy has been selected. Finally, I show how to extend all of these results to judgments among multiple alternatives through the kind of sequential balloting typical of the legislative amendment process.

Voting in Search of the Public Good: the Probabilistic Logic of Majority Judgments

1. Introduction

The defining feature of democratic government is its control by citizens through the votes of majorities. But there is no consensus among political theorists about precisely what benefits the ballot is supposed to confer upon democratic societies. The three most prominent views are:

1. Voting provides a way to aggregate the individual preferences of citizens into decisions that equitably represent the preferences of the society;
2. Voting provides a check on government power via the ability of voters to throw the rascals out;
3. Voting is a means by which a society may identify policies that best promote the public good.

These views need not be mutually exclusive. The ballot may serve each in different contexts, and may serve several at once. But political theorists often champion one conception as the primary advantage of democracy over other forms.¹ And they often deny the meaningfulness or usefulness of other conceptions, or argue that properly understood these other conceptions are subsumed under their favored view.

Consider the view of *preference aggregators*. With the ascendance of social choice theory it has been widely held that democratic voting is a means whereby a society may aggregate the preferences of individuals into an equitable expression of the preference of the group.² The preferences of individuals are usually presumed to reflect their private interests. Such notions as *the public good* or *the public interest* make sense, preference aggregators contend, only insofar as they may be reduced to a tally of the preferences of individuals. Any other notion of the public good is metaphysical nonsense.

Opponents of the preference aggregation conception of voting argue that the view is ultimately inappropriate or even incoherent as a theory of democratic government. They point to Arrow's Theorem and related results from social choice theory that arguably show that no voting system that satisfies very minimal requirements of fairness will invariably aggregate collections of individual preferences into fair decisions for the group. They point out, for instance, that simple majority preferences may well be cyclic. A majority of voters may prefer A to B, and a majority may prefer B to C, and at the same time a majority

may prefer C to A. Thus, any voting system that merely aggregates preferences may produce social choices that depend on such arbitrary matters as the order in which options come to a vote. (See Sen, 1982, Ch. 8, “Social Choice Theory: A Re-examination,” esp. pg. 163, where he forcefully argues that the use of the Method of Majority Decision is inappropriate for interest aggregation.)

Another contingent of theorists, the *power checkers*, maintain that since democratic voting is not a fair means of aggregating preferences, its only function is to keep government power in check (see Riker, 1982). *Power checkers* recognize that government provides certain necessary benefits and that no society can endure without it. But government institutions and officials tend to project their power to an extent that may become oppressive to the citizenry. The primary advantage of democracy over other forms is the restraint it places on the power of government by the ability of the ballot to overturn laws and remove officials from office.

Those who maintain that there is a public good that supersedes mere majority aggregates of individual preferences have a tougher row to hoe in defense of democracy than *preference aggregators* and *power checkers*. For, *public good seekers* are faced with two daunting tasks: they need to show that some notion of the public good has substance, the ontological problem; and they need to show how democratic institutions may come to identify and implement laws and policies that are among the best options for promoting the public good, the epistemological problem. They must at least provide grounds for thinking that democracy may in principle be as competent as other forms at discovering and implementing policies that promote the public good. Failing this, *public good seekers* who champion democracy must retreat to the view that in spite of its inferior ability to secure the public good, democracy confers compensatory benefits. If democracy cannot secure the public good to the same extent as enlightened autocracy, then democracy’s virtue must lie, e.g., in the checks it places on the potential abuses of power, to which autocracies are highly susceptible.

As daunting as the ontological and epistemological problems may seem to be, I think that both may be successfully met. First, with regard to the ontological problem, the difficulty is not that we lack plausible accounts of the public good. Several credible views are available. I will briefly describe some of them in

Section 2, accounts that derive from the views of Aristotle, Locke, Rousseau, Mill, and Rawls. So, with regard to the ontological problem the main difficulty is to determine which one most plausibly characterizes the kind of good that public policy should promote. I will not advocate a specific view, but will rest content with reminding the reader that at least some notions of the public good are not mere metaphysical moonshine.

Assuming that there is a public good for a given society, how can democratic institutions discover which among the alternative proposed laws or policies or candidates would best promote that good? This, the epistemological problem, is very large topic. I will only address one aspect of it in this paper. I will focus on the probabilistic logic of voting. I will explicate a model of democratic voting that shows how majorities may achieve an extremely high degree of competence at judging which among alternative policy options would best promote the public good. In particular I will present two extremely robust versions of the *Condorcet Jury Theorems*. These theorems show that if, on average, voters possess a very moderate degree of individual competence at recognizing which among pairs of policy proposals (or laws, or candidates) would best promote the public good, then, provided that each voter votes his or her own best judgment, the majority is extremely likely to select the better policy.³

The efficacy that Jury Theorems ascribe to democratic voting may fail to issue, but only because certain requisite conditions may be violated in practice. Voters may, for example, align themselves too much with factions and vote in blocks; and the average propensity for voters to discern the better of two policies or candidates might be too low. Thus, the Jury Theorems show how good democracy *could be* at identifying the public good, not how good it is in fact. But such *mere normativity* is no failing. The most that may be hoped of any political theory is that it describes what consequences may be expected if political institutions conform to a given model. Although all such models are idealizations, they may, nonetheless, help explain the behavior of real political institutions and suggest more effective ways to structure them.

Here is how I will proceed. In Section 2 I will briefly explore the ontological problem. I will survey several plausible conceptions of the public good that may ground an epistemic conception for at least some

kinds of issues that may be decided by democratic voting. My point here will not be to give anything like a definitive argument that there is a public good or that any one of these conceptions gets it right.

Furthermore, none of these conceptions need be wedded to the idea that *every* issue voted on is tied to the public good it identifies. Rather, my point is to illustrate that for a wide range of philosophically respectable views there is such a thing as the better policy in at least some cases, and to show that such views may find aid and comfort from what Jury Theorems imply about the ability of majorities to find the better policy.

In the remainder of the paper I will turn to the probabilistic logic of voting to see how it bears on the epistemological problem, the problem of ascertaining which of several alternative policy proposals may best promote the public good when there is one. Section 3 presents some technical and conceptual background needed for an understanding of the Jury Theorems. In particular I will present and discuss the notions of voter competence of probabilistic independence employed by these theorems.

In Section 4 I explicate two comprehensive Jury Theorems and their implications for public good seeking votes. My presentation is designed to provide the reader with a thorough understanding of these theorems. It presupposes only a passing acquaintance with probability and statistics. The formal results are described (and proved in an Appendix) from the ground up; so the reader may verify that no hidden assumptions have slipped by.

After the votes are counted and the majority's choice has been adopted, to what extent should those who voted for the alternative be persuaded that the better policy has won? Jury Theorems do not speak directly to this issue. In Section 5 I will present a *Convincing Majorities Theorem*, a form of Bayes' Theorem that shows that under conditions where Jury Theorems apply, the number of votes by which a proposal defeats an alternative may provide evidence that the winner is very probably the better option, evidence that should logically persuade even initially skeptical members of the minority.

The results of Sections 4 and 5 apply only to judgments between *pairs* of alternatives. In Section 6 I extend these results to multiple alternatives through the kind of pairwise sequential balloting that arises in the legislative amendment process.

The subject of Jury Theorems and their applicability to democratic voting has had a small but devoted following in recent years. But even among devotees there remains some disagreement about the applicability of these theorems to democratic voting. Most versions of Jury Theorems require *probabilistic independence* among voters, and it is often suggested that this is not at all feasible in a democracy, where public discussion is both a basic right and an indispensable part of the democratic process. Rawls argues that Jury Theorems are inapplicable on precisely these grounds (1971, p. 358), and even such devotees as Ladha (1992) and Grofman and Feld (1988) succumb on occasion.⁴ I will attempt to clarify this issue at several points in Section 3. However, the two Jury Theorems I will offer are more general than most and do not require probabilistic independence among voters.⁵ Rather these versions of the theorem show precisely how the group success rate at choosing the better policy increases as the *degree of independence* among voters increases. Thus, independent voting is desirable but not required for the theorems to apply.

Furthermore, I will argue that the kind of independence that is relevant to Jury Theorems is not at all the kind of restriction on voter communication some have taken it to be. Rather, in the voting context the criterion of *probabilistic independence* may be satisfied if after the public debate each voter assesses the merits and votes his or her own best judgment. I will argue that under reasonable conditions *probabilistically independent voting* may be accomplished by means of *independent voting* as the term is commonly understood—*voting one's conscience* rather than as part of a block or faction.

One of the strengths of Jury Theorems is the way they elucidate how certain characteristics of a democratic society may corrupt its ability to implement the best policies. At the same time, these theorems show that other characteristics are not as injurious to the pursuit of the public good as one might have thought. Factionalism and the slavish adherence to the views of opinion leaders are shown to be especially harmful. However, if such vices can be avoided, the majority will very likely succeed in identifying the best policy, even when average voter competence is only mediocre. But, the Jury Theorems are two-edged. If the average voter competence is too low, they cut the opposite way; the probability becomes high that the majority will select an inferior alternative. Thus, no advocate of democracy who holds there to be a public good in at least some policy domains, for at least some issues, can afford to ignore the implications

of these theorems.

2. Several Conceptions of the Public Good

Theories of government often appeal to some notion of the public good. In this section I describe several accounts of the kind of public good that government should promote, those of Aristotle, Locke, Rousseau, Mill, and Rawls. Each view is well known. So I will present only enough detail to remind the reader that there are plausible conceptions of the public good. Each view gives substance to the idea that some public policies may better promote the public good than others and that voters may reasonably attempt to discern the better policy rather than vote their private interests.

2.1 Aristotle. Aristotle's account of the public good emerges in the *Politics* (1971b) and the *Nicomachean Ethics* (1971a). In the *Politics* he explains the purpose of the state as follows:

... a state is not a mere society, having a common place, established for the prevention of mutual crime and for the sake of exchange. These are conditions without which a state cannot exist; but all of them together do not constitute a state, which is a community of families and aggregations of families in well-being, for the sake of a perfect and self-sufficing life. ... The end of the state is the good life, and these are the means towards it. And the state is the union of families and villages in a perfect and self-sufficing life, by which we mean a happy and honorable life. (1280^b-1281^a) ...the form of government is best in which every man ...can act best and live happily. (1324^a)

So Aristotle maintains that the purpose of government is to provide the requisite conditions for a life of happiness and well-being for all. Presumably government is supposed to enact whatever policies and laws will most likely promote this kind of public good.

To get a clearer picture of the kind of public good that Aristotle has in mind we need to get a better handle on what he means by a life of happiness and well-being. In the *Nicomachean Ethics* he investigates the concept of happiness and proposes a connection between happiness and human virtue. He first says that happiness is a final good and the chief good. It is desirable only for its own sake, whereas honor, pleasure, and every other virtue are desirable both for themselves and for the sake of happiness (1097^a-1097^b). Aristotle then delves more deeply into the nature of happiness.

... to say that happiness is the chief good seems a platitude, and a clearer account of what it is still desired. This might perhaps be given, if we could first ascertain the function of man. ... Now if the function of man is an activity of soul which follows or implies a rational principle, ... human good turns out to be activity of the soul in accordance with virtue, and if there are more than one virtue, in accordance with the best and most complete. But we must add 'in a complete life'. (1098^a)

What does it mean for human good to be activity of the soul which follows a rational principle in accordance with virtue? The idea seems to be that human beings are endowed with natural biological and intellectual needs and desires, and are naturally disposed to attempt to fulfill them. These natural needs and desires are not necessarily the same as what people consciously want; rather, they are those aspirations that are most deeply rooted in human nature. Virtue is the skill to act in the way best suited to bring about the fulfillment of these natural needs and desires. The good (happy) life is, then, a life of skillful action in pursuit of the fulfillment of those needs and desires that are most deeply rooted in human nature.

Aristotle goes on to say that chance plays an important role in human happiness—that circumstance and ill fortune can crush and maim happiness for even the most virtuous of men. So, both a virtuous character and moderately favorable circumstances are required for a good human life. Presumably, then, on an Aristotelian account of the public good, government should attempt to enact those laws and policies that will most likely provide circumstances favorable for all members of society to pursue and fulfill those needs and desires that are most deeply rooted in human nature. Thus, on an Aristotelian conception, when offered a pair of policies (or candidates for office), each citizen should vote for the one that in his or her judgment will provide conditions most conducive to a fulfilling human life for all members of society.⁶

2.2 Locke. In the opening section of the *Second Treatise on Civil Government* (1685) John Locke describes the purpose of government to be, “for the regulating and preserving of property, and of employing the force of the community, in the execution of such laws, and in the defense of the commonwealth from foreign injury; and all this only for the public good.” (I,3) Later, in a section called “Of the Ends of Political Society and Government,” Locke further asserts that men unite under commonwealths in order to mutually protect their *property*, by which he means their lives, liberties and

estates (IX,123-124). This section concludes:

... the power of the society or legislative constituted by them can never be supposed to extend farther than the common good, but is obliged to secure everyone's property against those three defects above-mentioned that made the state of nature so unsafe and uneasy. And so, whoever has the legislative or supreme power of any commonwealth, is bound to govern by established standing law And all this to be directed to no other end but the peace, safety, and public good of the people. (IX, 131)

So, on Locke's view it is the responsibility of government to promote the kind of public good that comes from the establishment of peace and safety and the protection of life, liberty, and private possessions.

Locke holds that these goods all flow from a person's natural right, indeed obligation, to self preservation. He argues that each person is obliged to his creator to preserve his or her own life and, to the extent that it does not conflict with his or her own preservation, to preserve the lives of all other persons. Health, liberty, and possessions are the instruments that tend to preserve life (II, 6); and peace and safety are essential to these. On a Lockean view, then, voters and legislators in a democracy should vote for whatever policy (or candidate) will in their judgment best promote the conditions of peace, safety, health, liberty, and protection of private property necessary for all persons to flourish.

2.3 Rousseau. Contemporary proponents of an epistemic conception of voting as a means to promote the public good often find Rousseau's *Social Contract* (1762) congenial. Their regard for Rousseau seems to derive mainly from his distinction between *the will of all*, which is the aggregate of individuals' private interests, and *the general will*, which aims at the public good (see Bk. II, iv). This looks like just the right sort of distinction to set the epistemic conception of voting apart from preference aggregation views. But, although Rousseau seems to make the right distinction, he says precious little about the nature of the public good at which the general will aims. The following passages are among the few in *Social Contract* that describe the nature of the public good.

What is the goal set before themselves by all political organizations? -- surely it is the maintenance and the prosperity of their members. (Bk. III, ix)

So long as a number of men assembled together regard themselves as forming a single body, they have

but one will, which is concerned with their common preservation and with the well-being of all. When this is so, ... the State ... is not encumbered with confused or conflicting interests. The common good is everywhere plainly in evidence and needs only good sense to be perceived. (Bk. IV, i)

For Rousseau, then, the public good lies in promoting the prosperity and well-being of members of society.

The final sentence of the last passage seems a bit too simplistic. Although the public good may sometimes be plainly evident, the public policy option that would best promote it is often far from obvious. For example, the good health of all is clearly a public good; but it may be very difficult to discern whether a private health care system or some variety of government sponsored health insurance will best promote this good. However, proponents of the epistemic conception of voting have a remedy that seems to maintain the spirit of Rousseau's view. They appeal to Jury Theorems to show that majorities may, nevertheless, be highly successful at discerning which policy option will best promote the public good.⁷ A further passage from *Social Contract* is especially congenial to the epistemic conception of voting:

When a law is proposed in the assembly of the People, what they are asked is not whether they approve or reject the proposal in question, but whether it is or is not in conformity with the general will, which is *their* will. It is on this point that the citizen expresses his opinion when he records his vote, and from the counting of the votes proceeds the declaration of the general will. When, therefore, a view which is at odds with my own wins the day, it proves only that I was deceived, and that what I took to be the general will was no such thing. Had my own opinion won, I should have done something quite other than I wished to do, and in that case I should not have been free. (Bk. IV,ii)

In Section 5 I show how Rousseau's claim that the votes of the majority should rationally persuade the minority that they were mistaken, and the better policy has been adopted, may indeed be born out.

2.4 Mill. Preference aggregators sometimes regard utilitarianism as congenial. When citizens vote their personal preferences, the majority get what they desire, and happiness is maximized. However, John Stuart Mill is completely antagonistic to this view of voting. In *Utilitarianism* Mill contends that the good consists wholly in the greatest happiness in quantity and quality for all humankind. But in *On Liberty* and in *Considerations on Representative Government* Mill argues that government's proper role in promoting

happiness is to provide a public good that is quite distinct from enforcing the individual preferences of majorities. Mill maintains that promoting the public good is the chief responsibility of government, but holds that this is best achieved through the creation and maintenance of conditions under which each person may enjoy the widest possible scope to pursue his or her own course to happiness. In *On Liberty* he supports this view by arguing that no person can reliably determine the best course to happiness for others and that more happiness will be available to each person if he or she is permitted to pursue it with as little interference as possible (see esp. Ch. 1 and 4). So to promote the greatest amount of happiness, the best course for government is to provide conditions that permit and enable the widest possible scope for each person to pursue his or her own path.

Given these views about the proper role of government, what does Mill see as the role of the ballot? In *Considerations on Representative Government* he contends that each person is morally obliged to vote for those public policies that in his or her best judgment will most likely promote the public good. And Mill specifically argues that citizens should not vote their individual preferences or attempt to promote their personal advantage.

In any political election ... the voter is under an absolute moral obligation to consider the interest of the public, not his private advantage, and give his vote, to the best of his judgment, exactly as he would be bound to do if he were the sole voter, and the election depended upon him alone. (Ch. 10)

Mill supports this view by arguing that voting is not properly understood as a right; no one can have a right to power over other people. Rather, the vote is properly understood as a public trust. A person's vote, "... is not a thing in which he has an option; it has no more to do with his personal wishes than the verdict of a jurymen. It is strictly a matter of duty; he is bound to give it according to his best and most conscientious opinion of the public good. Whoever has any other idea of it is unfit to have the suffrage" (Ch. 10)

2.5 Rawls. In *A Theory of Justice* John Rawls (1971) identifies justice as, "the first virtue of social institutions, ... laws and institutions no matter how efficient and well-arranged must be reformed or abolished if they are unjust." (p.3) For Rawls the paramount public good is the establishment of institutions and laws that sustain a just society. So a Rawlsian conception of the public good may be

derived from his conception of justice.

Rawls argues that for a society to be just it must satisfy the following two *principles of justice*:

- (1) Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberties for all.
- (2) Social and economic inequalities are to be arranged so that they are both: (a) to the greatest benefit of the least advantaged ..., and (b) attached to offices and positions open to all under conditions of fair equality of opportunity. (see pp. 302-303)

Rawls adds to these principles *priority rules* that address trade-offs among liberties and social and economic advantage. Roughly, liberties may only be abridged to enhance the system of liberties for all, and only when acceptable to those with less liberty, but never in exchange for increases in social or economic advantage; also, any inequality of opportunity must enhance the opportunity of the least advantaged.

Ultimately Rawls takes his specific principles and rules to be special cases of a general conception of justice that may guide the supplementation or modification of the specific rules if they prove inadequate.

All social primary goods--liberty and opportunity, income and wealth, and the bases of self-respect--are to be distributed equally unless an unequal distribution...is to the advantage of the least favored. (p. 303)

However, for Rawls the just society is not an end in itself. Rather, it provides the requisite conditions for citizens to have reasonable prospects of achieving their individual *rational plans of life* (pp. 92-93). People may choose distinct rational plans of life with distinct ends. But, whatever their ends, certain primary goods are necessary means to successfully pursue them. So, the Rawlsian conception of the public good for which government is responsible consists in the creation and promotion of conditions conducive to a fulfilling life for each citizen in pursuit of his or her own plan of life; and the best instrumental means to achieve this good is to implement those policies, laws, and institutions that will most likely bring about the equitable distribution of liberty and opportunity, income and wealth, and the bases of self-respect.

2.6 The Nature of the Public Good. Each of the political theories summarized here maintains that there is a public good that lies within the province of government. Each maintains that at least some issues that

come to a vote are concerned with finding the policy that would best promote that good. Furthermore, these theories agree to a remarkable extent about what is most fundamental to the nature of the public good: it consists in the enactment of policies and laws that provide conditions under which citizens may best pursue good lives. These views mainly diverge with regard to the principles governments should follow as the best instrumental means to secure this good (e.g. the protection of property, the preservation of liberty, the equitable distribution of goods, etc.). But, whatever the view, a fundamental issue remains: is a democratic society capable of using the ballot to choose the best public policies? Jury Theorems answer this question with a qualified *yes*.

In the next section I describe two Jury Theorems. I will state these theorems in terms of “voting to achieve *the* public good.” However, these theorems apply equally to any specific conception of the public good. That is, as stated these theorems say that if the average propensity of voters to identify which of two policies (or candidates) will best promote *the public good* is a little above $\frac{1}{2}$, and if the number of voters is large enough, and if each voter tends to vote his or her own best judgment, then the majority will very probably choose the better policy. However, the theorems apply equally if, for example, the term ‘Rawlsian public good’ is uniformly substituted for the term ‘public good’. (A policy better promotes the Rawlsian public good than an alternative if it is better at fostering the Rawlsian just society—i.e., it better promotes extensive, equal basic liberties, while only permitting significant social and economic inequalities that yield compensatory benefits for the least favored.) Thus, *if* the average propensity of citizens to identify which of two policies will better promote *the Rawlsian public good* is a little above $\frac{1}{2}$, and if the number of voters is large enough, and if each voter tends to vote his or her own best judgment, then the majority will very probability choose the policy that will better promote *the Rawlsian public good*. These theorems apply similarly to the Aristotelian conception, the Lockean conception, utilitarian conceptions, or any other coherent view. So, for the remainder of this inquiry precisely which conception of the public good is correct or appropriate for a given society will be of no particular import.

For our purposes it will not usually matter whether voters have any very clear conception of the nature of the public good. Nor will it matter whether voters consciously attempt to enact the better policy. Their

propensities to vote for the better option may be accidental or a side-effect of some other aim. The Jury Theorems apply even if voters are mindless idiots. However, if voters possess at least some vague impression of the nature of the public good and deliberately attempt to discern the better policy, we have better reason to suppose that their propensities to succeed are, on average, better than chance; and this is crucial if the majority is to have a reasonable prospect of choosing correctly. So I will speak in terms of voters who *pursue* the public good. But I will not generally assume that voters have a clear or principled conception of its nature.

Throughout the remaining inquiry I will speak in terms of *the* public good (for a given society). But readers should keep the caveats of the previous two paragraphs in mind. The central issue, then, is this: to what degree is it likely that the best policies for promoting the public good will come to be adopted by the votes of majorities? The answer given by the Jury Theorems is that, whether or not any particular *theory* of the public good gets it right, the *ballot* will almost surely do so, provided only that the average propensity for individuals to perceive the better of two policies is a little better than random chance and each votes his or her own best judgment (and not in alliance with a block or faction). For practical purposes this answer should certainly be good enough. For, whether or not some *theory* of the nature of the public good gets it right, in a democracy it is ultimately the voters and legislators who have the responsibility to find the best means to bring about conditions under which citizens may have a reasonable chance of living good lives. To the extent that they may succeed, they can only do so through successive attempts to enact practical laws and policies to achieve that good, issue by issue and vote by vote.

3. Dim Perception and Judgmental Competence

In Book III of the *Politics* Aristotle suggests that deliberation by the multitude may provide an advantage in determining the good of the state.

The principle that the multitude ought to be supreme rather than the few best is one that ... seems to contain an element of truth. For the many ... when they meet together may very likely be better than the few good, if regarded not individually but collectively For each individual among the many has a share of virtue and prudence, and when they meet together, they become in a manner one man, who

has many feet and hands and senses; that is a figure of their mind and disposition. (1281^b).

In the same paragraph Aristotle concedes that this principle is not true of every deliberative body, since some bodies consist of men that differ little from brutes. But he immediately reaffirms that, "... there may be bodies of men about whom our statement is nevertheless true." A bit further on, after discussing whether the ballot should be reserved to those who have proper expert knowledge, he adds, "Yet possibly these objections are to a great extent met by our old answer, that if people are not utterly degraded, although individually they may be worse judges than those who have special knowledge—as a body they are as good or better." (1282^a, 15)

How are we to understand Aristotle's suggestion that the multitude are better judges of the good of the state than those with special knowledge? Clearly one consideration is that the multitude provides a greater resource of ideas and experience. The exchange of opinions supplements the limited knowledge of each and widens his or her perspective; it makes each more aware of the possible impact of a proposed policy on others. Public debate may sharpen ideas, uncover their implications, explore their strengths and expose their weaknesses more fully than the more parochial reflections of an elite few. This seems to be what Aristotle had in mind. But, granted all of this, is there any reason to think that, after the public discussion, when it comes to a vote, the multitude may *judge better* than the enlightened few?

The issue here is not the danger of political corruption. Both ruling elites and the multitude may be blinded by self-interest. The issue is whether a multitude of well-intentioned public good seekers may be superior to an enlightened elite at discerning the better policy. It may be extraordinarily difficult for even the most astute, fair-minded person to anticipate the consequences of policy options. Consider the issue of whether the government should implement a national health insurance system. There are well-informed, well-intentioned people on both sides of this issue who believe that nearly everyone would receive the greater benefit from the policy they support. Public policy is replete with such hard issues, where it is difficult to discern the costs, benefits, and future consequences of alternative policy options. Despite such difficulties we will see that the judgment of the multitude may indeed be superior to that of the elite few.

3.1 Individual Competencies and Average Individual Competence. Let us take as granted that a

primary function of democratic government is to implement policies and laws that promote the public good and that in at least some cases there is an objective (though unknown) fact of the matter as to which of several policies will best promote that good. Pairs of policy options are presented to voters or legislators, and each attempts to discern which option will most likely promote the public good, and votes accordingly. The policy options may, e.g., consist of two versions of a national health insurance system; or one option may be a specific public insurance system and the alternative may be to maintain a wholly private system. When more than two options are considered, they typically compete through a series of contests between pairs, as through the legislative amendment process. I will discuss multiple options and sequential binary contests in Section 6.

Think of each voter as a sort of limited, imperfect detector that attempts to discern which of two policies will better promote the public good. To function in this role, voters need have no very clear conception of the nature of the public good. They need only possess some propensity to correctly perceive the better of two policies through the influence of their experience and attention to the issues and concerns raised in public debate. Under these conditions it can be shown that if the number of voters is large, and if each votes his or her best judgment, then provided that the average *competence level* (i.e. *detection accuracy*) among individual voters is at least slightly better than chance and the voting population is sufficiently large, the majority vote is extremely likely to select the better policy. This is the main thrust of the Jury Theorems.

In Section 4 I will present two extremely robust Jury Theorems. But to fully understand these theorems the reader will need to understand a certain amount of technical, conceptual background. The subsections of the present section will explain these details. In doing so it will delve into the notion of probabilistic independence that is relevant to Jury Theorems.

Let us begin by representing the voting context more formally. Label the two alternative policy options ‘W’ and ‘U’. What the voter is trying to determine is whether it would be more in the public interest, all things considered, to adopt W or to adopt U. Let ‘B’ represent the proposition that W will produce a greater public good than U, all things considered; and let ‘C’ represent the contrary proposition, that U will

result in a greater public good than W.

The *competence level* of a voter i relative to proposition B is the probability, r_i , that he or she will vote that B if B is true (i.e. if W is a better policy than U). The *competence level* of voter i relative to B is a *conditional probability*, and will be represented as follows: $P[b_i=1 | B] = r_i$. The expression ‘ $b_i=1$ ’ is shorthand for the assertion that voter i casts 1 vote that B holds; alternatively, ‘ $b_i=0$ ’ represents the claim that voter i casts 0 votes that B. Technically b_i is called a *random variable*. I will assume each voter either votes for B or for the alternative claim, C. So, if $b_i=0$, then voter i must cast a vote for C and $c_i=1$; if $b_i=1$, then voter i doesn’t vote for C, so $c_i=0$. The expression ‘ $P[b_i=1 | B] = r_i$ ’ says that the probability that person i will cast his or her vote for B, if B is in fact true, is r_i ; and the expression ‘ $P[c_i=1 | C] = r_i$ ’ reads similarly. I will assume that each individual i has the same competence level for C (if true) as for B (if true)—i.e. that $P[b_i=1 | B] = P[c_i=1 | C] = r_i$. This assumption is not strictly necessary but the presentation would become somewhat more complicated without it. It follows from the rules of probability theory that $P[b_i=0 | B] = 1-r_i$ and $P[c_i=0 | C] = 1-r_i$. (See the Appendix for a summary of defined symbols.)

It may well be that no one has any clear idea of the competence level r_i of an individual voter i . Presumably it depends on such factors as education, astuteness, attention to the public debate, and life experience. Fortunately, *knowledge* of voters’ competence levels will turn out to be unnecessary.

Let n be the total number of voters and suppose that each voter i has a competence level r_i on the kind of issue at hand in the existing social environment. Individual competence levels at identifying the better policy will surely differ from voter to voter and from issue to issue. But, whatever their individual competencies, the *average competence level* of the whole group on the present kind of issue is some number $r = \sum_{i=1}^n r_i / n$. Only the average competence level of the group will play a significant role in what follows. Furthermore, the precise value of r will not much matter either, as we will see.

For a population of n voters, each voter i ultimately either votes for B (so that $b_i=1$) or votes for C (so $b_i=0$). The sum $\sum_{i=1}^n b_i$ represents the total number of votes for B. If we divide this sum by the total number of voters, we get the fraction of the votes in favor of B. Let ‘ $\%b$ ’ represent this fraction: $\%b = \sum_{i=1}^n b_i / n$. When $\%b > 1/2$ the majority have voted in favor of B; if $\%b < 1/2$, then $\%c > 1/2$, and the majority have voted

against B and for C. (For convenience I ignore the special case where $\%b = \frac{1}{2}$; if n is odd $\%b$ cannot be $\frac{1}{2}$; if n is even and not small it is extremely unlikely that $\%b$ will be precisely $\frac{1}{2}$. Although this case is easily handled, it would complicate the presentation unnecessarily.)

If a primary function of government is to implement policies that promote the public good, then it is vital that majorities select the better policy most of the time. That is, it is vital that, with high probability, over half the voters will cast votes for B when B is true. So we want $P[\%b > \frac{1}{2} \mid B]$ to be as close to 1 as possible; and similarly, we want $P[\%c > \frac{1}{2} \mid C]$ to be near 1. Just as the value r_i of $P[b_i = 1 \mid B]$ and $P[c_i = 1 \mid C]$ represents the individual competence level of voter i , the values of $P[\%b > \frac{1}{2} \mid B]$ and of $P[\%c > \frac{1}{2} \mid C]$ represent the *group competence level* of the population of voters. If both of these probabilities are close to 1, then whichever claim is correct, B or C (whichever policy is better, W or U), will very probably receive the majority of votes. The group competence levels $P[\%b > \frac{1}{2} \mid B]$ and $P[\%c > \frac{1}{2} \mid C]$ behave similarly; so with no loss of generality we may focus attention on $P[\%b > \frac{1}{2} \mid B]$.

The Jury Theorems will describe bounds on the values of the group competence levels $P[\%b > \frac{1}{2} \mid B]$ and $P[\%c > \frac{1}{2} \mid C]$. These bounds depend on just four factors: the number of voters, n , the average voter competence level, r , and two further factors that I will label ‘ s^2 ’ and ‘cov’. I explain these two factors next.

3.2 Variance from the Average Competence Level and Covariance Among Individual

Competencies. Consider two societies of equal size, each consisting of n voters, and each possessing the same average value of individual voter competencies, r , at discerning the better of two proposals. The way in which individual competencies are distributed among members of the group may be quite different for the two societies, and yet their average value, r , may be precisely the same. For example, suppose each society has 1000 voters. In one society every member has the same individual competence level, $r_i = .55$, so that the average value $r = .55$; and notice that for this society the majority has *some chance* of selecting an inferior alternative. The other society, let us suppose, consists of 550 voters who “always get it right” (i.e. have $r_i = 1$) and 450 voters who “always get it wrong” (i.e. have $r_i = 0$). This group, too, has average competence level $r = .55$, but the majority will *always* select the better alternative. Thus, the way in which individual voter competencies are spread out around their average value, r , affects the probability that the

majority will choose correctly. The factor s^2 , called the *variance in the distribution of voter competencies*, provides a measure of how widely the individual competencies are distributed around the group average. By definition $s^2 = (1/n) \sum_{i=1}^n (r_i - r)^2$; it is the average of the squared distances of individual competencies from the average individual competence level for the group. (The factor s , the square root of s^2 , is called the standard deviation of individual competencies about the average competence level.)

Notice that when all voters have the same competence level, the average value r for the group equals the individual competence level of each; in that case $s^2 = 0$. But generally there will be some diversity among individual competencies in a group; so $s^2 > 0$. For a given value of r , the maximum possible value of s^2 occurs when $r \cdot n$ of the voters “always get it right” and the remaining $(1-r) \cdot n$ voters “always get it wrong”. In this case $s^2 = r \cdot (1-r)$. We will see that the larger s^2 is, the better chance the majority has of choosing correctly. And, although it may be difficult to estimate the value of s^2 for an actual population of voters, it turns out that the precise value s^2 takes will be relatively unimportant.

The factor ‘cov’ represents the average *covariance* among voters; it is a measure of how independently they tend to vote. To define ‘cov’ we must consider *joint probabilities* of votes for policies. And, to get a good understanding of what cov means, we will need to delve more deeply into the nature of the propensities that these probabilities represent.

For any two voters i and j , let $r_{j,i}$ be the probability that both j and i will vote for B if B is true. Formally, $r_{j,i} = P[b_j=1 \cdot b_i=1 \mid B]$, whereas $r_j = P[b_j=1 \mid B]$. Think of these probabilities like this. Although it is natural to speak of ‘ $P[b_j=1 \mid B]$ ’ as representing j ’s propensity to vote for B (when B is true), this way of speaking is a bit misleading. The propensity more properly belongs to the whole system, the social and political context, of which j is a part—it is the propensity of that system, through its institutions, the public debate, and other relevant information and experiences it provides to j , to bring j to vote for B (when true). Indeed, the system affects all voters in a way that issues in propensities for complete profiles of votes to be cast. The propensity of the system to produce a complete profile of votes in which *voter 1* votes for B , *voter 2* votes against B , ..., and *voter n* votes for B is represented by the probability ‘ $P[b_1=1 \cdot b_2=0 \cdot \dots \cdot b_n=1 \mid B]$ ’. Each possible profile has some propensity of occurring; and the

propensities of all possible profiles must sum to 1. But different profiles may have widely different propensities. If, for instance, i and j agree to always vote the same way, the propensity of any profile in which j votes differently from i (e.g. $b_j=1$ and $b_i=0$) will be 0. Thus, how one person votes may be tightly bound up with how others vote, and this will show up in the propensities for various possible complete profiles of votes.

The propensity for j to vote correctly, $P[b_j=1 \mid B] = r_j$, is related to the propensities for complete profiles in a simple way. Its probability value is just the sum of the propensities of all complete profiles in which j votes for B (i.e. in which $b_j=1$). Now consider just those complete profiles in which j and i both vote for B . If we sum the propensity values of all such profiles, the result is $P[b_j=1 \cdot b_i=1 \mid B] = r_{j,i}$, which represents the propensity of the system to bring both j and i to vote for B . If j and i agree to always vote the same way, then the propensity for j and i to both vote for B (if true) is just that for j to vote for B , since i always votes with j : $P[b_j=1 \cdot b_i=1 \mid B] = P[b_j=1 \mid B] = P[b_i=1 \mid B]$ (i.e. $r_{j,i} = r_j = r_i$). If j and i always vote in opposition, then $P[b_j=1 \cdot b_i=1 \mid B] = r_{j,i} = 0$, regardless of the values of r_i and r_j individually. In both cases j 's and i 's votes are completely correlated. If j and i vote independently, their joint propensity to vote for B (when true) will just equal the product of their individual propensities: $P[b_j=1 \cdot b_i=1 \mid B] = P[b_j=1 \mid B] \cdot P[b_i=1 \mid B]$ (i.e. $r_{j,i} = r_j \cdot r_i$). So the degree to which j and i tend to vote in agreement may be measured by the difference $(r_{j,i} - r_j \cdot r_i)$.

By definition, $\text{cov} = [2/(n \cdot (n-1))] \cdot \sum_{i=1}^{n-1} \sum_{j=i+1}^n (r_{j,i} - r_j \cdot r_i)$. In a population of n voters the number of pairs of voters is $n \cdot (n-1)/2$. Thus, cov sums the measures of each voter's tendency to vote in agreement with others and divides by the total number of voter pairs to produce the average value of these correlation factors.

3.3 On Voter Independence. To get a better handle on what voter independence means, imagine that the public debate has come to a close and its time to vote. Individuals j and i formulate their decisions in the context of the whole social-political environment, including the existence and nature of relevant social and political institutions, information garnered from the public debate, and other relevant information and life experiences each may have. Then, j and i vote independently just if the propensity of this context to bring

them to jointly vote for B (when true) is no stronger or weaker than the product of its propensities to bring each to vote for B. That is, the fact that i decides to vote for B (when true) provides *no additional influence or information* that affects the propensity for j to vote for B. On the other hand, if the fact that i votes for B provides an *additional positive indicator* (beyond everything covered by the social context, the quality of the debate, and j 's other information and experiences) that j will vote for B, then j and i do not vote independently, and $r_{j,i} > r_j \cdot r_i$. And if i 's vote for B provides an *additional indicator* that j will *not* vote for B, then $r_{j,i} < r_j \cdot r_i$. Thus, the difference $(r_{j,i} - r_j \cdot r_i)$ provides a measure of the increase or decrease in the likelihood that j will vote for B (when true) that comes from factoring in the additional information that i also votes for B. This difference indicates a degree of *excess influence* on j 's vote, influence in excess of the perfectly legitimate role i may play in trying to convince j of the truth of B.

The mere fact that (as part of the relevant context) j and i belong to the same political party or organization does not automatically mean that they fail to vote independently of each other. If the context, including her political views, give j a tendency (a propensity r_j) to vote as she does based only on her own best judgment, so that *given her propensity to vote as she does*, her actual vote provides no additional evidence for how individual i will vote (given what his own individual propensity), then she votes independently of i . Such independent voting need not require that j not belong to a political party. Independent voters may affiliate with a party because their views tend to agree with those the party promotes. But insofar as one votes independently, the party does not influence them to vote against their own best judgments. Alternatively, if, as a party member, j tends to override her own best judgment to some degree, and has even a slight tendency to vote a party line, and if i also does this, then j and i will tend to vote together more often than their individual propensities would warrant. In that case, how i votes provides an additional indicator that alters the propensities for j 's vote, an indicator of *excess influence*.

If voters have no excess influence on one another, then each term $(r_{j,i} - r_j \cdot r_i)$ is 0, so $\text{cov} = 0$. When there is excess influence, some of these terms may be positive and others negative. Just as there may be some tendency to vote with their friends, there may also be a tendency to vote in opposition to foes. So, even when there is excess influence, the terms that make up cov may largely cancel and cov may be near 0.

One additional point: cov as specified so far only applies to proposition B (i.e. option W is better than option U). But what of the value of cov relative to claim C (i.e. that U is better than W)? Obviously the same analysis applies. Cov may be defined for C in terms of factors $s_i = P[c_i=1 | C]$ and $s_{j,i} = P[c_j=1 \cdot c_i=1 | C]$. If $r_i \neq s_i$ and $r_{j,i} \neq s_{j,i}$, then cov for B will not usually equal cov for C. However, I will henceforth assume that $r_i = s_i$ and that $r_{j,i} = s_{j,i}$, so that cov is the same for both C and B. For, although we could treat competence levels for B and C (and their cov values) separately, nothing essential would be gained and the presentation would become somewhat more complicated.

3.4 More on the Nature of Probabilistic Independence. The most common objection to Jury Theorem models of voting is that they seem to presuppose a kind of independence among voters that prohibits communication about issues; yet, public discussion is essential. The viability of the democratic process requires that voters communicate. The exchange of views supplements the limited knowledge of each, widens his perspective, and increases awareness of the possible impact of policies on others. Public debate sharpens ideas, explores their strengths and exposes their weaknesses. If communication among voters were forbidden, their individual competencies at discerning the better policy would surely suffer, and their average individual competency, r , might well fall below $\frac{1}{2}$, with devastating consequences for the group's chances of selecting the better policy (as we will see). But doesn't public discussion make the judgment of each voter dependent on that of others (making cov large)? Thus, Jury Theorems may seem to identify a devastating dilemma for democracy. Voters may secure suitable individual competencies only by making their judgments excessively interdependent, which may result in a low level of group competence, so that the better policy runs a high risk of defeat.

I have already attempted to counter this objection to some extent by suggesting that if group members vote their own best judgment's, communication among them does not make them probabilistically dependent. But this issue is so important that some additional clarification is in order.

First, the Jury Theorems presented here will not assume probabilistic independence. These theorems hold for a very general model of voting, a model that applies in any case where among the proposals there is a *best policy* to accomplish some goal. The theorems will show precisely how the average individual

competence level for the group may combine with a measure of the degree to which voting is (or is not) independent to produce the competence level of the group at choosing the better policy. Nevertheless, a significant implication of these theorems is that when the average individual competence level is at least slightly better than chance, the more independently people vote, the higher the probability that the majority will select the better policy. So a fairly high degree of independent voting remains very desirable.

It turns out that the kind of probabilistic independence at issue for Jury Theorems has very little to do with whether an exchange of views among members of the group influences the *formation* of their individual propensities to choose the better policy. The relevant kind of probabilistic independence may obtain even when there is a high degree of *propensity forming interaction* among individuals. A simple example will illustrate the point. Consider the following model. A thousand pennies are loosely piled, facing "heads up", in a small cardboard box. The box is placed on an anvil and struck by the large, irregularly shaped head of a machine-operated sledge hammer. As a result of the impact each penny becomes warped. The particular warp of a penny depends on such factors as its location in the box relative to the irregularities in the hammer's head, the points of contact between neighboring pennies in the pile, and the amount by which neighboring pennies transmit the force of the impact to bend their neighbors. Now, suppose that the warp of each penny affects its propensity to come up "heads" when tossed. So the propensity of a penny towards "heads" is *dependent* on its interactions with its neighbors. Nevertheless, the propensity of a given penny towards "heads" is *probabilistically independent* of the *outcomes of tosses* of the other pennies. If the warp of a given penny produces a propensity of .61 for "heads", then the likelihood that tossing it will produce "heads" is .61, regardless of how tosses of other pennies turn out. This kind of probabilistic independence is just the sort that is germane to Jury Theorems.

There is one sense in which the outcomes of tosses of other pennies may be probabilistically relevant to the outcome of a toss of *this penny*. That is, if *we don't know* the propensity of *this penny* for "heads", then outcomes of tosses of similarly bent neighbors may furnish us with evidence about *this penny's* propensity. (Indeed, *this penny's* propensity is most strongly evidenced by repeatedly tossing it; the Law of Large Numbers says that its frequency of "heads" on a long series of tosses will almost certainly

approximate its true propensity.) At any rate, *this penny* has some specific propensity for "heads" regardless of whether *we know* its value. And, relative to this propensity, the likelihood that a toss will produce "heads" is probabilistically independent of the outcomes of tosses of its neighbors. Relative to its true propensity, outcomes of tosses of other pennies provide absolutely no additional information about the likelihood that it will come up "heads" on a given toss.

Although the interaction among pennies in the bent-penny model is not precisely analogous to voter interactions, the analogy accurately reflects the relevant point about probabilistic independence. The location of a penny and its interaction with others under the impact of the sledge hammer shapes its individual propensity for "heads". Similarly, the individual experiences of a voter in conjunction with her assessment of the issues raised in the public debate shapes her propensity to come to a correct judgment when she casts her ballot. When *we don't know* her propensity to vote for a specific kind of policy, information about how others have decided to vote may be relevant to our assessment of how she may be inclined to vote. For, other voters also attempting to discern the better policy. But this sort of probabilistic relevance is perfectly consistent with the kind of *probabilistic independence* appealed to in Jury Theorems.

The treatment of voters as each having a certain propensity to vote for the better policy, given the social and political environment, is really not very different than treating each bent coin as though it has some propensity for "heads", given how coins are usually tossed. Of course, if enough details of a particular coin toss are spelled out – e.g. precisely how it leaves the fingers, with what forces, at what height above the ground, what air currents are present and how they interact with the coin, the nature of the ground's surface and the coin's impact with it – given all this, the outcome of a specific toss might well be deterministic, or might have quite a different propensity than that we would usually attribute to the coin *over to the range of usual circumstances*. So, when one models a coin as having a particular propensity for "heads", one is only proposing that there is a reasonable stochastic model of its behavior over the range of tosses of the usual kind. Similarly, when we model voters as each having a propensity to choose correctly, we are only proposing that over a range of circumstances of the relevant kind (i.e. for the kind of issue at hand, given the usual nature of the public debate and other information available to voters in such cases)

there is a reasonable stochastic model of their inclination to choose the better policy.

The mere existence of a reasonable stochastic model does not, of course, mean that anyone knows what it is. One may know that this bent coin has *some* propensity for *heads*, that *some* stochastic model must accurately reflect its behavior on tosses. But, at the same time, one may be completely in the dark about the numerical value of this propensity. Even so, laws of large numbers may assure one that, for any given propensity the coin may have, the coin's frequency of turning up *heads* is very likely to be near that propensity. Jury Theorems say roughly the same kind of thing about percentages of votes. They apply to groups of independent voters for any given profile of values for the individual's competence levels. And they apply regardless of whether anyone *knows* the profile of competencies for group members.

What kind of probabilistic dependence would cause trouble for the veracity of majority judgments as represented by Jury Theorems? In terms of the bent-penny model, suppose we instituted the following plan. We divide the group of 1000 pennies into 10 subgroups of 100. We select one penny from each subgroup, its "subgroup leader", toss it, and observe the result. If it's "heads", we place the other pennies from its subgroup in a vice and bend them enough to move their propensities more towards "head"; if it's "tails", we bend the other pennies from its subgroup enough to move their propensities more towards "tails". This process is repeated for each subgroup. Thus, the propensity of each penny becomes probabilistically dependent on its initial warp together with the "decision" of its subgroup leader.

This example illustrates the kind of probabilistic dependence that can undermine the high group competence levels promised by Jury Theorems, dependence that arises from *voting with a group* rather than *voting ones conscience*. In the context of Jury Theorems *probabilistically independent voting* may be accomplished simply by means *independent voting* as that notion is commonly understood. The legitimate purpose of the public debate is to help voters perceive more clearly the relative merits of policy options so that their individual propensities for selecting the better policy may improve. In that spirit it is perfectly legitimate for voters to attempt to influence one another, to get each other to recognize the perceived merits and defects of proposals. No harm (no increase in cov) results from that. But when a voter succumbs to *excessive influence* over his vote by paying too much heed to how others have decided to vote, he overrides

his own best judgment of the merits and fails to aid the group epistemologically. He merely amplifies the judgment of those he follows. Voting with a subgroup as a block (e.g. for the sake of solidarity) can undermine the ability of democratic government to identify and implement the best policies. However, if most voters are only swayed a little by such *excess influence*, then the degree of independence will remain high enough that cov will be fairly small. If only this could be achieved, then if the average individual competence level exceeds random chance by only a bit, the group competence level will excel, as the Jury Theorems will show.

Let me remind you once more that the Jury Theorems presented here will not rely on independent voting. But these theorems show that a fairly high degree of independent voting is often very desirable. After presenting the theorems we will we will briefly return to the topic of probabilistic independence (in Section 4.3) to see what the Jury Theorems say about the group competence level when the degree of voter independence is not particularly high.

4. Two Jury Theorems

We now turn to the presentation of two very powerful Jury Theorems. The first is a version of the *Weak Law of Large Numbers*. It presupposes nothing about how the group's votes tend to be probabilistically distributed. The second theorem supposes that the group's votes tend to be approximately normally distributed. This second theorem is closely associated with the *Central Limit Theorem*, which assures us that approximate normality is a very reasonable assumption in the contexts like the present one, where a "group probability function" derives from a large number of component probability functions.

4.1 A Weak Law of Large Numbers Version of the Jury Theorem. If a population of n voters casts precisely k votes for B , then the fraction of the votes cast for B is $\%b = k/n$. The term $P[\%b = k/n \mid B]$ represents the probability that B will get this fraction of the votes when B is true. If we multiply each possible value of $\%b$ by the probability that B will get that fraction, and then sum the resulting terms, we have what statisticians call the *expected value* of $\%b$ (given B): $r^* = \sum_{k=0}^n (k/n) \cdot P[\%b = k/n \mid B]$. Rather surprisingly, r^* must equal the average voter competence level, r . (See Appendix).

A closely related factor is the *variance of possible voting results* (possible values of $\%b$) about its

expected value r^* . It is customarily represented by the symbol ' σ^2 '. The formula $\sigma^2 = \sum_{k=0}^n ((k/n) - r^*)^2 \cdot$

$P[\%b = k/n \mid B]$ defines it. This variance provides a measure of how distant the most likely possible values of k/n tend to be from the *expected value* of the vote count, r^* . If σ^2 is small, the most likely results will be those with values near r^* ; large values of σ^2 indicate that a fair number of the more likely results have values rather far from r^* .

In the voting context as I've described it σ^2 turns out to be related to r , s^2 , and cov by the formula: $\sigma^2 = [r \cdot (1-r)/n] - [s^2/n] + ((n-1)/n) \cdot \text{cov} > 0$ (see Appendix). A little reflection on this formula shows this: whatever the value of r (the average voter competence level), and whatever the value of s^2 (the average squared distances of the individual competence levels from their average value r), if n is large and cov is small, then σ^2 must also be small. When σ^2 is small we get two important results. First, a form of the *Weak Law of Large Numbers* guarantees that the value of $\%b$ is very likely to be quite close to r . Second, another form of the *Weak Law of Large Numbers* guarantees that, *if* r is even a bit greater than $1/2$, *then* the majority will very probably vote for the better policy—i.e. the value of $P[\%b > 1/2 \mid B]$ is very near 1. *If*, on the other hand, the value of r is even a little below $1/2$, *then* the majority will very probably vote against the better policy—i.e. the value of $P[\%b < 1/2 \mid B]$ is near 1. The following theorem expresses these relationships.

Theorem 1. *Weak Law Jury Theorem.* (Proved in the Appendix.)

For n voters, with r , s^2 , and cov as defined above, $\sigma^2 = [r \cdot (1-r)/n] - [s^2/n] + ((n-1)/n) \cdot \text{cov}$, and:

$$(1) \quad \text{For all } \varepsilon > 0, P[-\varepsilon < \%b - r < \varepsilon \mid B] \geq 1 - \sigma^2/\varepsilon^2.$$

$$(2.1) \quad \text{If } r > 1/2, \text{ then } P[\%b > 1/2 \mid B] \geq 1 / (1 + \sigma^2/(r-1/2)^2).$$

$$(2.2) \quad \text{If } r < 1/2, \text{ then } P[\%b < 1/2 \mid B] \geq 1 / (1 + \sigma^2/(r-1/2)^2).$$

If $\text{cov} \leq s^2/(n-1)$, then, where R is defined by the formula $R = (1/[2 \cdot (r-1/2)]^2) - 1$:

$$(1') \quad \text{For all } \varepsilon > 0, P[-\varepsilon < \%b - r < \varepsilon \mid B] \geq 1 - [r \cdot (1-r)]/(\varepsilon^2 \cdot n).$$

$$(2.1') \quad \text{If } r > 1/2, P[\%b > 1/2 \mid B] \geq 1 / (1 + [r \cdot (1-r)]/[(r-1/2)^2 \cdot n]) = 1 / (1 + R/n).$$

$$(2.2') \quad \text{If } r < 1/2, P[\%b < 1/2 \mid B] \geq 1 / (1 + [r \cdot (1-r)]/[(r-1/2)^2 \cdot n]) = 1 / (1 + R/n).$$

Notice that if n is large enough, the term ' R/n ' in Theorem 1 approaches 0. So for large voting populations one of the probabilities $P[\%b > 1/2 \mid B]$ or $P[\%b < 1/2 \mid B]$ approaches 1, depending on whether $r > 1/2$ or $r <$

$\frac{1}{2}$. In either case the fraction of votes in favor of the better policy will very probably be near the average voter competence level r (clause (1')). Clauses (1') and (2.1') together tell us that when r is close to $\frac{1}{2}$ the majority is unlikely to be a *large majority* even when it is highly probable that the majority will choose correctly. Theorem 1 is a minimal result in the sense that for most voting populations the probabilities will be much higher than the lower bounds specified on the right-hand sides of the inequalities. Much better estimates of the kind of values these probabilities will take for most voting populations are provided in the next subsection.

Table 1 illustrates, for various values of r (when $\text{cov} = s^2/(n-1)$), the size of the voting population needed to insure that a given level of $P[\%b > \frac{1}{2} | B]$ is reached. If cov is smaller, the same value for $P[\%b > \frac{1}{2} | B]$ is achieved with smaller values of n ; if cov is larger, a larger value of n may be required. Each row shows, for a fixed value of r , the number of voters needed to obtain $P[\%b > \frac{1}{2} | B] = .999$, or $.99$, etc. Notice that the required population size drops off sharply as r rises above $.51$, and also as $P[\%b > \frac{1}{2} | B]$ recedes from $.999$. The table may also be used to see how values of r below $\frac{1}{2}$ affect $P[\%b < \frac{1}{2} | B]$. Take $' .999'$, $' .99'$, etc. as values of $P[\%b < \frac{1}{2} | B]$ but replace each value of r with $(1-r)$ (i.e. replace $.51$, $.52$, etc. with $.49$, $.48$, etc.).

Table 1

Group size n sufficient to make the group competence level $P[\%b > \frac{1}{2} | B]$ exceed p when the average individual competence level is r (no matter how $\%b$ is distributed), and $\text{cov} \leq s^2/(n-1)$.

$\backslash p $.999	.99	.98	.97	.96	.95	.90	.85	.80
$r \backslash $	-----								
.51	2,496,501	247,401	122,451	80,801	59,976	47,481	22,491	14,161	9,996
.52	623,376	61,776	30,576	20,176	14,976	11,856	5,616	3,536	2,496
.53	276,501	27,401	13,562	8,949	6,643	5,259	2,491	1,568	1,107
.54	155,095	15,370	7,607	5,020	3,726	2,950	1,397	880	621
.55	98,901	9,801	4,851	3,201	2,376	1,881	891	561	396
.60	23,976	2,376	1,176	776	576	456	216	136	96
.65	10,101	1,001	495	327	243	192	91	57	40
.70	5,245	520	257	170	126	100	47	30	21
.75	2,997	297	147	97	72	57	27	17	12

Versions of the *Law of Large Numbers* often presuppose probabilistic independence. Full independence would mean that the way in which other people vote is irrelevant to the propensity of voter i to vote for B. But regardless of whether independence may reasonably be expected of voters seeking the public good, Theorem 1 does not assume it. Rather, Theorem 1 employs cov as a measure of the degree to which voters vote independently. Thus, Theorem 1 provides a completely general model of how the degree of voter independence combines with the average voter competence level to yield a lower bound on the competence level of the group.

4.2 The Normal Jury Theorem. Theorem 1 makes no assumptions about how the possible values of $\%b$ are *probabilistically distributed*. That is, suppose we know the competence level r_i of every voter i , and suppose we also know the binary joint probabilities $r_{j,i}$ for pairs of voters. This information by itself fails to uniquely specify the probabilities that B will win by specific amounts (i.e. specific values for $\%b$). Two different groups of n voters with precisely the same individual voter competencies and the same pairwise joint probabilities need not share the same group likelihoods, $P[\%b = q \mid B]$, of casting q votes for B; and so they may not share the same group competence levels $P[\%b > \frac{1}{2} \mid B]$. Theorem 1 merely specifies a generous lower bound on the value of $P[\%b > \frac{1}{2} \mid B]$, a lower bound that remains in force for every group, no matter what its *distribution* of $\%b$. Theorem 1 is a *worst case* result that admits the whole range of possible distributions for $\%b$. But many of these would be very unnatural as models of the voting dispositions of human beings in search of the best public policies. Most statistically distributed attributes in large natural populations—e.g. height, weight, daily calorie intake, grade point average, rates of movie attendance on Sunday afternoons—are much better statistically behaved than many of the distributions permitted by Theorem 1. The distributions of most interesting attributes in large natural populations closely approximate the bell-shaped curve of a *Normal distribution*.

The lower bounds placed on group competence levels by Theorem 1 are boosted significantly when the possible values for voting results, $\%b$, are Normally distributed. Given the voting situation as I've described it, the claim that, when B is true, the distribution of $\%b$ is (approximately) Normal comes to just this: (1) the most likely value of $\%b$ is (very nearly) the average competence level r ; (2) as values of q

recede from r the likelihood that $\%b$ will equal q when the votes are counted falls off in the same way that the height of a bell-shaped curve falls off from its peak. Here condition (1), that the most likely value of $\%b$ is r , does not mean that $\%b$ is at all likely to take the value r , but only that r is more likely to result from the vote count than any other value. Indeed, the distribution of *the number of heads* on 100 tosses of a fair coin is very nearly Normal, but the likelihood that 100 tosses will produce exactly 50 heads is only about .08.

The assumption that $\%b$ is approximately Normally distributed is quite a weak one. The distribution of $\%b$ is tied to a large population of individual propensities; and such composites are almost always approximately Normally distributed—the greater the number of component distributions, the more nearly Normal the composite tends to be. So called *Central Limit Theorems* attest to this fact. I won't rehearse their details. It will suffice to note that they support the following claims:

- (1) If voting is independent and all voters possess the same competence level, r , $\%b$ has a *binomial distribution*. Then, for populations of $n = 11$ or more, the values of $P[\%b = m/n \mid B]$ closely follow a Normal distribution with mean r and variance $\sigma^2 = r \cdot (1-r)/n$. Jury Theorems are often stated in terms of binomial distributions; but the assumptions they employ are stronger than needed to obtain similar results.
- (2) If voting is independent but competencies are distinct, with average r , the variance σ^2 is $[r \cdot (1-r)/n] - [s^2/n]$. As n increases the distribution of $\%b$ will very closely approximate a Normal distribution.
- (3) A more general version of the Central Limit Theorem applies even when competencies are not independent. It says that if each voter in a large population is nearly pairwise independent of *most* other voters (i.e. if for each voter j , values of $(r_{j,i} - r_j r_i)$ are nearly 0 for all but a fairly small percentage of other voters i), then the distribution of $\%b$ approximates a Normal distribution (mean r , variance σ^2 as defined earlier).
- (4) *Central Limit Theorems* provide sufficient conditions for distributions to approach Normality, but not necessary conditions. Approximate Normality is ubiquitous for distributions of attributes in large

natural populations, even when interactions among members may appear to make them highly interdependent.

So, for large populations the distribution of %b may well be nearly Normal.⁸ And, of course, any set of assumptions that imply that the distribution of %b is nearly Normal must be at least as strong as the supposition that the distribution of %b is nearly Normal. Thus the following theorem, although less sweeping than Theorem 1, is more general than most other Jury Theorems.

Theorem 2. *Normal Jury Theorem.*⁹

For n voters, with r , s^2 , and cov as defined earlier, the mean value of %b (given B) is r and the variance must be $\sigma^2 = [r \cdot (1-r)/n] - [s^2/n] + ((n-1)/n) \cdot \text{cov}$, as proved for Theorem 1. If the distribution of %b, given B, is Normal (or approximately Normal), then the following assertions hold (or hold to a high degree of approximation). Here, by definition $N[z] = (2\pi)^{-1/2} \int_{-\infty}^z \exp[-x^2/2] dx$, which is the area under the *Standard Normal* bell-curve from $-\infty$ up to the point z on the x -axis.

$$(1) \quad \text{For all } \varepsilon > 0, P[-\varepsilon < \%b - r < \varepsilon \mid B] = 1 - 2 \cdot N[-\varepsilon/\sigma].$$

$$(2.1) \quad \text{If } r > 1/2, \text{ then } P[\%b > 1/2 \mid B] = N[(r-1/2)/\sigma].$$

$$(2.2) \quad \text{If } r < 1/2, \text{ then } P[\%b < 1/2 \mid B] = N[(r-1/2)/\sigma].$$

If $\text{cov} \leq s^2/(n-1)$, then, where R is defined by the formula $R = (1/[2 \cdot (r-1/2)]^2) - 1$:

$$(1') \quad \text{For all } \varepsilon > 0, P[-\varepsilon < \%b - r < \varepsilon \mid B] \geq 1 - 2 \cdot N[-\varepsilon/(r \cdot (1-r)/n)^{1/2}].$$

$$(2.1') \quad \text{If } r > 1/2, P[\%b > 1/2 \mid B] \geq N[(r-1/2)/(r \cdot (1-r)/n)^{1/2}] = N[(n/R)^{1/2}].$$

$$(2.2') \quad \text{If } r < 1/2, P[\%b < 1/2 \mid B] \geq N[(r-1/2)/(r \cdot (1-r)/n)^{1/2}] = N[(n/R)^{1/2}].$$

For specific values of r and n , the value of $N[(n/R)^{1/2}]$ may be found by calculating $z = (n/R)^{1/2}$ and looking up the probability value of $N[z]$ on any table for the *Standard Normal distribution*. Table 2 shows the population size needed to insure that specific probability values for $P[\%b > 1/2 \mid B]$ will be reached when $\text{cov} = s^2/(n-1)$. If cov is smaller, the same value for $P[\%b > 1/2 \mid B]$ is achieved with smaller values of n ; if cov is larger, a larger value of n is needed. Comparing Table 2 with Table 1 we see that when the distribution of %b is approximately Normal, the same high group competence levels may be achieved with much smaller voting populations than required by some *worst case distributions* encompassed by Theorem

1. Table 2 may also be used to see how values of r below $\frac{1}{2}$ affect $P[\%b < \frac{1}{2} | B]$. Just read ‘.999’, ‘.99’, etc. as values for $P[\%b < \frac{1}{2} | B]$ and replace each specified value of r with $(1-r)$.

Table 2
Group size n sufficient to make the group competence level, $P[\%b > \frac{1}{2} | B]$, exceed p when the average individual competence level is r , $\%b$ is Normally distributed, and $\text{cov} \leq s^2/(n-1)$.

$\backslash p $ $r \backslash $.999	.99	.98	.97	.96	.95	.90	.85	.80
.51	23,864	13,524	10,541	8,840	7,659	6,761	4,104	2,684	1,770
.52	5,959	3,377	2,632	2,207	1,913	1,688	1,025	670	442
.53	2,643	1,498	1,167	979	848	749	455	297	196
.54	1,483	840	655	549	476	420	255	167	110
.55	945	536	418	350	303	268	163	106	70
.60	229	130	101	85	74	65	39	26	17
.65	97	55	43	36	31	27	17	11	7
.70	50	28	22	19	16	14	9	7	5
.75	29	16	13	11	9	9	5	3	3

Theorem 2 does not presuppose independent voting. It employs cov to model the degree of voter interdependence. If voter competencies are independent, even just pairwise independent, then $\text{cov} = 0$.

However, cov may remain fairly small even when voting is somewhat inter-dependent. And when cov is small the effect of large group size on group competence is most dramatic. But, in any case, Theorem 2 supplies a general model of how individual competencies affect group competence, a model that explicitly take the degree of voter independence into account, whatever it may be.

4.3 Moderately Sized Values of Cov. When voters vote independently, their average covariance level, cov , will be 0. Then a *Central Limit Theorem* guarantees that the approximate Normality assumption of Jury Theorem 2 is satisfied. But even for groups of non-independent voters, approximate Normality may well hold. If cov remains fairly small a very general version of the *Central Limit Theorem* guarantees this. Even when cov is not so small, near Normality is ubiquitous among distributions on large populations. And, in the event that approximate Normality fails in some cases, the most general version of the Jury Theorem, Theorem 1, still applies. Thus, although the effects of large population size on group competence

levels is most striking when cov is small, similar benefits may result when cov is moderately large.

For an illustration of the effect of moderately sized values of cov, suppose (for simplicity) that each voter's individual competence level has the same value, r . In that case $s^2 = 0$. Furthermore, suppose that for each voter j , the positive and negative correlations with all but m other voters balance out to 0, and there only remains a *surplus positive correlation* with those m other voters. Now, keeping things simple, suppose that each voter j also has the same amount of positive correlation with each of these m other voters. The term in cov that represents the correlation of voter j with voter i is $(r_{j,i} - r^2)$, since $r_i = r_j = r$. And for this to be the same amount of correlation for all m other voters, $r_{j,i}$ must be a constant; call it 'c'. Thus, the factors accounting for the surplus correlation all have value $(r_{j,i} - r_j \cdot r_i) = (c - r^2)$.

The largest that c could possibly be is r . It has this value when j and i are as strongly correlated as they can be—i.e. $c = r_{j,i} = r_j = r$. We are assuming that c represents a positive correlation, so the smallest c can possibly be is r^2 , the value it would have if j and i were independent—i.e. $c = r_{j,i} = r_j \cdot r_i = r^2$. Now define $q = (c - r^2)/(r - r^2)$, the fractional amount $(c - r^2)$ is of the largest possible value that it could have. That is, the value that the correlation factor $(c - r^2)$ actually has is q times its greatest possible value: $(c - r^2) = q \cdot (r - r^2) = q \cdot r \cdot (1 - r)$.

There are n voters and each has a surplus correlation $(c - r^2)$ with m other voters. So there are a total of $n \cdot m$ such correlation terms. However, when you count terms this way, each term gets double-counted, since whenever i is correlated with j , j is also correlated with i . So there are actually only $(n \cdot m)/2$ such correlation terms in the formula for cov. We've been supposing that all other terms in cov balance out to 0.

$$\text{Thus, cov} = [2/(n \cdot (n-1))] \cdot \sum_{i=1}^{n-1} \sum_{j=i+1}^n (r_{j,i} - r_j \cdot r_i) = [2/(n \cdot (n-1))] \cdot [n \cdot m/2] \cdot (c - r^2) = [m/(n-1)] \cdot q \cdot (r - r^2).$$

$$\text{Therefore, } \sigma^2 = r \cdot (1-r)/n + [(n-1)/n] \cdot \text{cov} = (1 + q \cdot m) \cdot r \cdot (1-r)/n.$$

This equation for σ^2 , resulting from our simplified model, shows the effect that positive correlation among voters may typically have. The number of voters, m , with which each has a *surplus positive correlation* is between 0 and $n-1$. And the fractional value q of the amount of possible correlation between pairs of these voters lies between 0 and 1. At the extreme of total correlation, $m = n-1$ and $q = 1$, we have $\sigma^2 = r \cdot (1-r)$, which is the same as the variance for a single voter with competence level r . If, on the other hand, $m = 0$ or $q = 0$, then there is no net positive correlation and $\sigma^2 = r \cdot (1-r)/n$, which is the variance for n

completely independent voters. For an intermediate case, suppose each voter has a net surplus correlation with 18 other voters ($m = 18$), and suppose the strength of these net correlations is half of the maximum possible amount ($q = 1/2$). Then $\sigma^2 = 10 \cdot r \cdot (r-1)/n$. So, for this group of voters to attain the same level of group competence as enjoyed by a similar group of independent (or $\text{cov} = 0$) voters it would have to be 10 times larger than the size of the independent (or $\text{cov} = 0$) group. And the effect of $m = 18$ and $q = 1/2$ is precisely the same as that of $m = 9$ and $q = 1$, or $m = 36$ and $q = .25$, or $m = 72$ and $q = .125$, etc. The effect in each case is to degrade the group competence level to that of one-tenth as many independent (or $\text{cov} = 0$) voters. The effect is the same as it would be if a society consisted of $n/10$ voting blocks of 10 voters each, where each block votes univocally.

Thus, although the effects of large population size on group competence is most striking when voters are independent, similar beneficial effects may result when voters are far from independent. If the average covariance among voters is small (e.g. $\text{cov} \leq s^2/(n-1)$), the group competence levels will be on the order of those given in Table 2. However, if cov is somewhat larger, the effect is merely to lower the group competence level to that enjoyed by a somewhat smaller group of more nearly independent voters.

4.4 Implications. Let us pause to consider the nature of the model of voting we have been exploring. The Jury Theorem model is no different in principle from other stochastic models of natural phenomena. For instance, the propensity of a particular moose to reproduce in a given year depends on a complex combination of its individual traits, the characteristics of the herd, and numerous environmental factors. So it may be extremely difficult for population biologists to estimate with any degree of accuracy the reproductive propensity of a single individual. However, based on the size of the population, information about its general health, the abundance of food in its habitat that year, the harshness of the winter, and other such factors, it may not be overly difficult for the biologist to infer parameter values for a population model that estimates, though roughly, the likely rate of reproductive success for the whole population. The biologist need have no detailed knowledge of the characteristics of the individuals that make up the population. A rough model of how certain factors tend to affect the average reproductive success of the population as a whole may suffice.

Similarly, although we may not know the individual propensities of particular voters to discern the better policy (on a given kind of issue under a given set of circumstances), we may nevertheless be able to estimate very roughly the value of the *average competence level*, r , of a large population of voters (on such issues under such circumstances), given their average level of education, the nature of the issue under consideration, the quality of the public debate, and other relevant properties of the social context. The precise value of r is not critical. It need only be a bit above $\frac{1}{2}$. The larger the population, the closer to $\frac{1}{2}$ may r be while still producing an extremely high group competence level. And, although citizens of a large, complex society may have neither the time nor the expertise required to deal with the intricacies of the many policy issues that arise, they may well be sufficiently competent to generate a high group competence level at selecting a body of representatives whose average competence level at selecting the better policy is sufficiently high.

This assessment of the implications of Jury Theorems may seem overly optimistic. David Estlund (1993) raises a significant problem in this regard. He argues that Jury Theorems cannot supply the kind of epistemic justification of democratic voting some have sought, because the average individual competence level of voters cannot be known. Estlund's article is primarily directed at showing how difficult it is to provide an epistemic justification of authoritarianism—i.e., how difficult it is to try to show that the wisest and most politically skilled should rule. But he directs his critique at the epistemic justification of democracy as well. With regard to rule by the wise, Estlund points out that if citizens are to accept authoritarian rule on epistemic grounds, their recognition of the legitimacy of their rulers will crucially depend on the ability to know that their rulers are indeed politically wise. But citizens have no way of knowing this unless they can independently confirm that the policies their rulers have enacted are for the best. And, Estlund argues, they have no good way to do this. Estlund then contends that democracy is in a similar bind. Referring explicitly to Jury Theorems, he argues that citizens have no good way of coming to know whether the average individual competence level is sufficiently greater than $\frac{1}{2}$: "... it is hard to see how such a thing could be established without independent public knowledge of the answer key—the very facts we hope to use democratic voting to reveal." (p. 93) Apparently, equal experts are likely to disagree

about whether the chosen policies are indeed among the best on offer; so we get little help from them. And we cannot measure voter competence by past performance, since in the political sphere it is often difficult to determine whether the results of an implemented policy are better or worse than those that would have issued from an alternative.

My response to Estlund has two parts. First, I will point out that Jury Theorems have important implications for democratic voting, regardless of whether anyone *knows* the average competence level of voters. This response does not dispute Estlund's main point, which is that Jury Theorems fail to provide citizens with an *epistemic justification* of democracy. My second response will attempt to take on Estlund's objection directly. I will argue that under the right circumstances it may indeed be possible for citizens to come to have good reason to believe that the average competence level is sufficiently high.

First, then, even if a society has no good means of assessing the average competence levels of voters and legislators, the Jury Theorem model still applies. Its veracity does not depend at all on whether anyone is *aware* of the model or *knows* the values of its parameters. Jury Theorems show that for a sufficiently large population of fairly independent voters, if the average voter competence level r is in fact a bit above $\frac{1}{2}$, then the majority will very probably choose the better alternative, regardless of whether anyone knows its value. Similarly, if r is a bit below $\frac{1}{2}$, the majority will very probably choose the inferior alternative. And a lack of voter independence will not tend to improve the situation. To the extent that votes are positively correlated (i.e. cov is large), the result is, in effect, the same as reducing the population size—for r above $\frac{1}{2}$, the majority becomes correspondingly less likely to choose wisely; for r below $\frac{1}{2}$ the majority is correspondingly less likely to choose poorly. In any case, although the difficulty in assessing average competence may undermine the ability of Jury Theorems to provide citizens with an *epistemic justification* of democracy, nevertheless, so long as the society remains committed to democracy (for whatever reason), it is vital to the well-being of citizens that it attempt to foster in them and their legislative representatives high enough competence levels at perceiving the public good to bring the average suitably above $\frac{1}{2}$, even if there is no good way to *measure* it.

As an illustration of the impact of voter competence, consider a group consisting of 350,000 voters,

about the number registered to vote in an average U.S. congressional district. Suppose this electorate has an average competence level of at least .51 at identifying the more competent candidate for office. And suppose they tend to vote their own best judgments (i.e. cov is small). Then, according to Theorem 1, the majority will have a better than .99 chance of choosing the more competent candidate, regardless of the underlying distribution over possible vote totals (i.e. over $\%b$). And, if the distribution over possible vote totals is approximately Normal, then Theorem 2 says that the chance of the majority choosing the more competent candidate is better than .9999. Furthermore, if on a given kind of policy issue a body of 435 elected representatives has an average competence level of .55 or better at discerning which of two proposals will best promote the public good, then Theorem 1 says that if they vote their consciences, the group competence level will be above .81; and Theorem 2 says that if the distribution over possible vote totals is nearly Normal, then the competence level of the legislature must be more than .98. However, if the average competence level in each of these cases is below $\frac{1}{2}$ by a corresponding amount, then the inferior candidate or policy is likely to be chosen. E.g., if the average competence level of the 435 representatives is below .45, then Theorem 1 says that the probability that the majority will vote for the inferior proposal is above .81; and Theorem 2 says that for a Normal distribution the chance of adopting the inferior policy must be at least .98. So the group competence level either suffers severely or benefits immensely from the average individual competence level, regardless of whether anyone *knows* its numerical value. Thus, Jury Theorems demonstrate the paramount importance of fostering institutions that are likely to effect a better than even rate of average voter competence, even if its value cannot be measured.

Secondly, despite the difficulties Estlund raises, I think that under the right circumstances citizens may indeed be able to ascertain that the average competence level is high enough, commensurate with group size, to supply a high group competence level. Admittedly, the prospects of this seem dim from our present perspective, as citizens of democracies as they now operate. Contemporary democratic societies are far too factious. So, even if the competence of citizens and legislators at recognizing the better policy is, on average, sufficiently high, our ability to discern it is greatly attenuated by the political environment. Voters and legislators too often set aside their own best judgments in order to maintain solidarity with their

ideological comrades. Legislators must do so to receive the kind of support they need from their parties and interest groups to gain a reasonable chance of re-election. And because voters tend to follow their parties' lines, the kind of deliberative, honest reflection that might signal competent discernment is not in evidence. Thus, circumstances as they now exist may not be conducive to a sound assessment of average competence levels. But average competence may well be more easily ascertainable in a different political environment.

In what sort of political environment might citizens have the ability to ascertain whether that average competence of voters and legislators is sufficiently above $\frac{1}{2}$? Consider a society with the features and institutions of the well-ordered society described by Rawls (1971, esp. Ch VIII): "...one designed to advance the good of its members and effectively regulated by a public conception of justice. ...in which everyone accepts and knows that the others accept the same principles of justice, and the basic social institutions are known to satisfy these principles. ...its members have a strong and normally effective desire to act as the principles of justice require." (pp.454-455). Presumably the citizens of this society are well educated and well informed, and understand in broad outline the kind of the public good that government should promote (e.g. that it calls for the enactment of policies and laws that provide conditions under which citizens may best pursue happy and fulfilling lives, as explicated in Section 2). For such a society the only issue before voters and legislators is which policy or candidates for public office will most likely produce this effect. Let us suppose in addition that the society is non-factional in the sense that, although proposals may be hotly debated, political parties and interest groups do not command allegiance to a party line. Rather, the reigning ethic is that, after fully attending to the debate, when it comes time to vote, each is to vote his or her own best judgment in an honest attempt to further the good of all. Adherence to this ethic would be evidenced by the fact that any given pair of voters may vote on opposite sides of issues from time to time. Under such conditions citizens would clearly have good reason to think that voters are fairly independent.

In such a society, if the average competence level is sufficiently high, it may be fairly apparent that it is. For, since citizens and legislators truly desire the good of all, if after a reasonable period of time the adopted policy appears to be inadequate, they should willingly reconsider. This environment should be

conducive to experimentation—e.g. the funding of test programs to explore approaches. And policy results may be compared with those of alternatives implemented by other governments. All of this may contribute to a long historical record for the society that indicates the degree to which policies have succeeded. And although the interpretation of this record may be controversial in various cases, sober non-partisan reflection may well provide a fair indication of the rate of policy success. Thus, the citizens of such a society may well come to have good reason to think that the average competence level of voters and legislators has come to be sufficiently greater than chance.

The ability to overcome factiousness and partisanship and to pursue the good of all over narrow self-interest may or may not be compatible with human nature. This is an empirical issue. But clearly some political philosophers have thought it possible. In the note at the beginning of *Social Contract* Rousseau says he will be considering “human beings as they are and laws as they might be.” So, given his description at the beginning of Book IV, of upright and simple men deciding affairs of state for the common good, Rousseau must think that our natures are capable of it. Those who champion an epistemic conception of the ballot mainly disagree with Rousseau with regard to how plainly evident the better policy would be among such citizens. Rather, they contend that only a better than chance level of voter competence need be evident, so that citizens may to be convinced that the majority will choose rightly.

If Rousseau’s vision seems overly utopian, consider some other views. In *Considerations on Representative Government* Mill forcefully argues that voters *should* rise above self-interest and vote their own best judgments of what is in the public interest. He clearly thinks that people are capable of it. And in *A Theory of Justice* Rawls goes to some length (in Chapter VIII) to “... sketch the development of the sense of justice as it presumably would take place once just institutions are established and recognized to be just.” (p. 453) He makes the case that human beings have the capacity to institute and maintain a well-ordered society as he conceives it. Through an examination of the relevant principles of moral psychology, he attempts to show that his conception is psychologically suited to human inclinations. And, although Rawls does not directly address the issue of factionalism, the way he argues for the psychological plausibility of the character traits needed by citizens of the well-ordered society provides a model of how

this might be done. I haven't the space to pursue this further here. But, I hope to have at least convinced you that we should not cynically dismiss the possibility that factionalism may be diminished, and that citizens and legislators may come to use the ballot in pursuit of the public good, and that their competence at doing so may become evident enough.

5. To What Degree Should the Majority Vote *Convince* Members of the Minority?

Although the veracity of Jury Theorem models of voting depends not at all on whether anyone knows the value of the average competence level, r , or the degree of voter independence, let us now suppose that voters are aware of these theorems and are confident that r is at least a bit above $\frac{1}{2}$ and that cov is small. When the balloting is over, to what extent may the vote count provide evidence that the better proposal has been adopted? Although Jury Theorems are often interpreted as providing an answer to this question, they do not speak to it directly. Jury Theorems only say how likely it is that the majority will vote for B (that W is better than U) *if* B is true. They do not address the inverse question: how likely is it that B is true *if* the majority votes for it? There is a connection between these two sorts of probabilities; they are related by the formula called Bayes' Theorem. In this section we will see what Bayes' Theorem says about the degree to which the vote count should rationally convince citizens that the winning proposal is the better alternative.

5.1 Prior and Posterior Probabilities, and the Size of a *Convincing Majority*. Consider the case of a *probabilistically rational voter* α . To say that α is *probabilistically rational* just means that she reasons in accord with the logic of probabilities. That is, there is a probability function ' P_α ' that represents α 's rational degrees of confidence in various propositions related to policy proposals and vote tallies. For example, however confident α may be in a proposition B (that proposal W is better than U), her degree of confidence in the alternative claim C (that U is better than W) must be correspondingly high or low—i.e. if $P_\alpha[B \text{ or } C] = 1$ (if α is certain that either B or C), then $P_\alpha[C] = 1 - P_\alpha[B]$. And in a similar vein, probabilistic rationality implies that for any relevant proposition E, the sum of α 's degree of confidence that both B and E are true and her degree of confidence that both C and E are true should just equal her degree of confidence in the truth of E—i.e. if $P_\alpha[B \text{ or } C] = 1$, then $P_\alpha[B \cdot E] + P_\alpha[C \cdot E] = P_\alpha[E]$.

Probabilistic rationality provides for conditional degrees of confidence as well, degrees of confidence

that may arise when new evidence is considered. If E is evidence that bears on the truth of B, α 's degree of confidence in B upon learning that E should satisfy the formula $P_\alpha[B | E] = P_\alpha[B \cdot E] / P_\alpha[E]$. Similarly, the likelihood that E will in fact occur when B is true satisfies the formula $P_\alpha[E | B] = P_\alpha[B \cdot E] / P_\alpha[B]$. From these two formulas it follows that α 's degree of confidence in B upon learning E should be related to the likelihood that E will occur if B is true by a form of Bayes' Theorem: $P_\alpha[B | E] = P_\alpha[E | B] \cdot P_\alpha[B] / P_\alpha[E]$. From this and the other constraints on probabilistic rationality already mentioned, another form of Bayes' Theorem follows: $P_\alpha[B | E] = P_\alpha[E | B] \cdot P_\alpha[B] / (P_\alpha[E | B] \cdot P_\alpha[B] + P_\alpha[E | C] \cdot P_\alpha[C])$. Dividing both the numerator and denominator of this formula by the numerator and recognizing that $P_\alpha[C] = 1 - P_\alpha[B]$ yields the following more compact form: $P_\alpha[B | E] = 1 / (1 + (Q \cdot R))$, where $Q = (1 / P_\alpha[B]) - 1$ and $R = P_\alpha[E | C] / P_\alpha[E | B]$. This form of Bayes' Theorem will be employed in a moment.¹⁰

For our purposes the import of Bayes' Theorem is this. Suppose that when it comes time to cast her vote, α believes that C is very probably true (that proposal U would better promote the public good than W) and votes accordingly. Her (low) degree of confidence in the truth of B when she casts her vote, after the public debate but *prior* to the vote count, is her *prior probability* for B, $P_\alpha[B]$. Once the results of the balloting are in, α may take the vote count as new evidence, E, and revise her degrees of confidence in B and in C. Her new rational degree of confidence in B, based on the information E, is called her *posterior probability* for B, her *posterior degree of confidence* in B. Bayes' Theorem expresses the relationship between a probabilistically rational person prior and posterior degrees of confidence.

Probabilistic rationality has surprising implications for how the evidence provided by the vote count should transform α 's prior degree of confidence into her posterior degree of confidence in B. Suppose that α 's prior degree of confidence is only $P_\alpha[B] = .05$. And suppose that she knows that %b is approximately Normally distributed and is quite confident that $r \geq .55$ and cov is very small. Then Bayes' Theorem says that *logically* a vote count consisting of $m = 519$ votes for B from $n = 1000$ voters should transform α 's prior probability into a posterior probability of $P_\alpha[B | \%b = 519/1000] \geq .99$ -- i.e. α should come to be at least 99% certain that B is true. This result follows from the next theorem. It implies that if probabilistically rational voters have good reason to believe that the average individual competence level is

at least a little above $\frac{1}{2}$ and that cov is small, then a rather small majority in favor of one option over another should *logically* convince even those who voted with the minority that the better option has very probably been chosen.

Theorem 3. *Convincing Majorities* Theorem. (Proved in the Appendix.)

For n voters, with r , s^2 , and cov as defined earlier, the mean values of %b (given B) and %c (given C) are r and the variance of each is $\sigma^2 = [r \cdot (1-r)/n] - [s^2/n] + ((n-1)/n) \cdot \text{cov}$, as proved in Theorem 1. If the distributions of %b (given B) and %c (given C) are (approximately) Normal, then (for $\exp[x] = e^x$, $e \approx 2.718$, the base of the natural logarithm):

$$(1) P[\%b = m/n \mid B] \approx \exp[-(m/n - r)^2/2\sigma^2] / (2\pi)^{1/2} \cdot \sigma \cdot n, \text{ and}$$

$$P[\%b = m/n \mid C] = P[\%c = (n-m)/n \mid C] \approx \exp[-((n-m)/n - r)^2/2\sigma^2] / (2\pi)^{1/2} \cdot \sigma \cdot n.$$

Where $k = m - (n-m)$ is the absolute number by which the votes for B exceed those for C:

$$(2) P[\%b = m/n \mid C] / P[\%b = m/n \mid B] \approx \exp[-k \cdot (r - 1/2) / (n \cdot \sigma^2)] = 1 / \exp[k \cdot (r - 1/2) / (n \cdot \sigma^2)], \text{ so}$$

$$(3) P_a[B \mid \%b = m/n] \approx 1 / (1 + (Q \cdot R)), \text{ where } Q = (1 / P_a[B]) - 1 \text{ and } R = 1 / \exp[k \cdot (r - 1/2) / (n \cdot \sigma^2)].$$

If $\text{cov} \leq s^2/(n-1)$, then:

$$(2') P[\%b = m/n \mid C] / P[\%b = m/n \mid B] \leq 1 / \exp[k \cdot (r - 1/2) / r \cdot (1-r)], \text{ so}$$

$$(3') P_a[B \mid \%b = m/n] \geq 1 / (1 + (Q \cdot R^*)), \text{ where } Q = (1 / P_a[B]) - 1 \text{ and } R^* = 1 / \exp[k \cdot (r - 1/2) / (r \cdot (1-r))].$$

Theorem 3 shows that when cov is small the size of the majority *fraction* of votes, %b, is not really the important factor in determining the posterior probabilities. Rather, the crucial factor is the *excess number* of votes B receives beyond the number of votes for C. That is, let '#b' and '#c' represent the number of votes for B and for C, respectively; so #b - #c = m - (n - m) = k. Then for cov small enough that $(n-1) \cdot \text{cov} \leq s^2$ we have that $n \cdot \sigma^2 \leq r \cdot (1-r)$, which yields (2') and (3') from (2) and (3). In that case the absolute number of voters, n , becomes irrelevant. Only #b - #c = k and r play a significant role. Thus, if cov is small, $P_a[B \mid \%b = m/n]$ is effectively just $P_a[B \mid \#b - \#c = k]$.

Table 3 shows, for values of r between .51 and .75 (and $\text{cov} \leq s^2/(n-1)$) and for values of $P_a[B]$ between .01 and .90, the differences #b - #c = k between numbers of votes for and against B that would suffice to produce a posterior probability for B of at least .99. (I've selected a posterior of .99 here just for

the sake illustration.)

Table 3

Difference $k = m - (n - m)$ between the number of votes for and against B sufficient to transform a prior probability $P_a[B] = p$ into a posterior probability $P_a[B \mid \%b = m/n] \geq .99$ when the average individual competence level is r , $\%b$ is Normally distributed, and $cov \leq s^2/(n-1)$.

$\backslash p \mid$.01	.02	.03	.04	.05	.10	.15	.20	.30	.40	.50	.60	.70	.80	.90
$r \backslash \mid$	-----														
.51	230	212	202	194	188	170	158	149	136	125	115	105	94	80	60
.52	115	106	101	97	94	85	79	75	68	62	57	52	47	40	30
.53	76	70	67	65	63	56	53	50	45	42	38	35	31	27	20
.54	57	53	50	48	47	42	39	37	34	31	29	26	23	20	15
.55	45	42	40	38	37	34	31	30	27	25	23	21	19	16	12
.60	22	20	19	19	18	16	15	14	13	12	11	10	9	8	6
.65	14	13	12	12	11	10	10	9	8	8	7	6	6	5	4
.70	10	9	8	8	8	7	7	6	6	5	5	4	4	3	3
.75	7	6	6	6	6	5	5	4	4	4	3	3	3	2	2

Inspection of Table 3 shows that even when a voter is quite uncertain about the precise value of r , but confident that it is *at least* some specific small amount greater than $\frac{1}{2}$ (and if she is confident that cov is small), then even a relatively few “excess votes” for B should be quite convincing. The entry in Table 3 under the column headed ‘.05’, for instance, indicates that when the average competence level r is .51, a vote count of 188 more votes for proposition B than against it is enough to transform α ’s prior probability $P_a[B] = .05$ to a posterior probability $P_a[B \mid \%b = m/n] = P_a[B \mid \#b - \#c = 188] = .99$. This depends only on the difference in the number of votes for and against B, i.e. only on the fact that $\#b - \#c = 188$. The values m and n make no difference. Similarly, if $r = .52$, then 94 more votes for B than against it will raise a prior of .05 to a posterior of .99; if $r = .53$, then a 63 vote difference will do the same. Hence, the likelihood of obtaining a convincing majority (e.g. at the .99 level, from a prior of .05) is the likelihood of obtaining an absolute vote difference for B appropriate to the value of r . Under the assumptions of Theorem 3 the size of the voting population plays no direct role in *how convincing* the outcome should be. Only the average competence level and the absolute number by which votes for B exceed those against can affect how convincing the results of the tally should be.

Notice that the *fraction* of votes required to produce a convincing majority decreases towards $\frac{1}{2}$ as the population size increases. I.e., for a fixed vote difference k , the absolute number of votes B receives must be $m = (k+n)/2$, since $k = m - (n-m)$. The fraction of votes for B that corresponds to this is $m/n = (\frac{1}{2} \cdot k/n) + \frac{1}{2}$, which goes to $\frac{1}{2}$ as n increases.

A natural reaction to Theorem 3 may be that although it is mathematically correct, intuitively it is just wrong; α 's opinion about the truth of B should not be influenced by the majority B at least not so strongly influenced by such small majorities: she has formed her opinion on the basis of a lifetime of personal experience and information gathered from attention to the public debate; other people's opinions, as expressed through their votes, should not undermine her convictions. This reaction expresses a misunderstanding of the voting situation *as it should be* when the group is faced with the vagaries and uncertainties that attend a search for policies that promote the public good. In this context the ballot is not a battleground where competing views clash for dominance, but a cooperative effort. Differences in perception about which policy would be for the best are like perceptual differences among a group of climbers navigating through a fog. It benefits no one to get his way if it leads them all into the abyss. Rather, each member attempts to perceive and evaluate the relative merits of the available courses of action from his own location and experience, from what he can discern by his own lights and infer from the reports of others. On this model, voting is a means of aggregating the individual perceptions of each into a composite estimate of the best course of action. It is in the interest of each to realize that his own perception is limited and may be illusory, that the composite judgment of the group may well be superior.

Concession to the better judgment of a group is quite reasonable if average individual competence is bit better than chance (commensurate with group size) and if systematic error is avoided by keeping cov small. However, even a very high posterior probability for the superiority of a given policy should leave some room for doubt. Policies should be reconsidered and confronted with new alternatives from time to time. Indeed, the high posterior probability for a policy is based only on whatever evidence was available before balloting together with the resulting vote count. So, although confidence in a new policy may start out high immediately after the votes are counted, the performance of the policy over time will surely

provide powerful additional evidence for or against its effectiveness at promoting the public good. Progress in the sciences depends on the continual reappraisal of even the most successful hypotheses, and on the re-evaluation and re-working of apparently unlikely alternatives. Clearly, policies intended to promote the public good deserve as much critical scrutiny and re-evaluation. However meritorious a particular policy may seem to be at a given moment, circumstances may change or a better alternative may be discovered.¹¹

5.2 The Likelihood of Obtaining a *Convincing Majority*. How likely is it that a *convincing majority* of votes will result from the balloting? We just saw that when *cov* is small, only the absolute number of votes by which B wins over C matters. The population size, *n*, makes no difference. However, the size of the voting population does have an important affect on posterior probabilities in the following sense: it affects the likelihood that a *convincing majority* of votes for B will be forthcoming. Theorem 4 makes this explicit.

Theorem 4. The Likelihood of Obtaining a *Convincing Majority*.

For *n* voters, with *r*, *s*², and *cov* as defined previously, the mean values of %*b* (given B) and of %*c* (given C) are *r* and the variances must be $\sigma^2 = [r \cdot (1-r)/n] - [s^2/n] + ((n-1)/n) \cdot \text{cov}$ (as proved in Theorem 1). If the distributions of %*b* given B and of %*c* given C are (approximately) Normal, then the following assertions about the probability of obtaining a majority of size *m* or greater hold (to close approximation). Here again, *k* = *m* - (*n* - *m*) is the number of votes for B in excess of those for C:

$$P[\%b \geq m/n \mid B] = N[(r - m/n) / \sigma] = N[(r - 1/2) - 1/2 \cdot k/n) / \sigma].$$

Also, if $\text{cov} \leq s^2/(n-1)$, then we have the following relationships:

$$\begin{aligned} P[\%b \geq m/n \mid B] &\geq N[(r - m/n) / (r \cdot (1-r)/n)^{1/2}] = N[(r - 1/2) - 1/2 \cdot k/n) / (r \cdot (1-r)/n)^{1/2}] \\ &= N[(n/R)^{1/2} - 1/2 \cdot k/(n \cdot r \cdot (1-r))^{1/2}], \text{ where } R = (1/[2 \cdot (r - 1/2)]^2) - 1. \end{aligned}$$

Theorem 4 follows directly from the basic properties of the Normal distribution. See, e.g., (Feller, 1968).

Notice that even when *cov* is small, the size of the voting population, *n*, remains crucial to the values of the right-hand sides of the equations—the absolute difference in votes, #*b* - #*c* = *k*, does not suffice alone, as it does in Theorem 3. Compare the clauses of Theorem 4 to clauses (2.1) and (2.1') of Theorem 2, and notice how similar they are. Indeed, the factor $1/2 \cdot k/n$ (and $1/2 \cdot k/(n \cdot r \cdot (1-r))^{1/2}$) in Theorem 4 goes to 0 as *n*

increases. So, comparing Theorems 4 and 2 shows that as the population size becomes large, the probability of achieving a convincing majority approaches the probability of achieving a simple majority, and both approach 1.

Table 4

Group size n sufficient for the likelihood of obtaining a *convincing majority*, $P[m-(n-m) \geq k | B]$, to exceed p when the average individual competence level is r , $\%b$ is Normally distributed, and $cov \leq s^2/(n-1)$. Here a *convincing majority* consists of at least k more votes for B than against B , where a k vote difference is sufficient to transform a prior probability $P_a[B] = .05$ into a posterior probability $P_a[B | \%b = m/n] \geq .99$.

Note: for $\%b$ Normally distributed and $cov \leq s^2/(n-1)$, k is uniquely determined by the value of r .

r	$\%p$.999	.99	.98	.97	.96	.95	.90	.85	.80
.51	188	40,482	29,310	25,933	23,951	22,539	21,440	17,994	15,942	14,459
.52	94	10,113	7,323	6,480	5,985	5,632	5,357	4,497	3,984	3,614
.53	63	4,498	3,260	2,885	2,665	2,509	2,387	2,005	1,777	1,612
.54	47	2,521	1,826	1,616	1,493	1,405	1,337	1,122	995	903
.55	37	1,600	1,158	1,024	945	890	846	710	629	570
.60	18	388	281	249	230	216	206	172	153	139
.65	11	162	117	103	95	89	85	71	63	57
.70	8	85	62	55	51	48	45	38	34	31
.75	6	50	36	32	30	28	27	23	20	18

Table 4 is very much like Table 2. Whereas Table 2 specifies, for values of r , the population size n that will with probability p produce a *majority* for B (if B is true), Table 4 specifies the population size n that will with probability p produce a *convincing majority* for B . Here, for illustration, a *convincing majority* is taken to be a majority large enough to transform a prior probability of .05 to a posterior probability of .99. These majorities correspond to the vote differences for B represented by the column under ‘.05’ in Table 3. This illustrates the way in which the Jury Theorem model of voting implies that a large voting population will *very probably* produce majorities large enough to convince even initially very skeptical, but probabilistically rational members of the minority that the better policy has been adopted.

6. Extensions to Multiple Alternative Proposals

The Jury Theorems and related results presented in Sections 4 and 5 apply only to choices between pairs of alternatives. But in searching for the best policy we must often consider a host of alternative

proposals. In this section I extend the previous results to judgments among multiple alternatives.

6.1 Extensions of the Jury Theorems to Multiple Alternatives. The Jury Theorems may be applied to cases involving a number of alternative proposals by running a series of pairwise contests. Suppose a series of h proposals, labeled V_1, V_2, \dots, V_h , vie for implementation as the policy that will, all things considered, best promote the public good. Two natural methods for conducting contests among the alternatives are: (1) conduct an *exhaustive series* of binary contests wherein each proposal stands in direct competition with each alternative; (2) conduct a *one-pass series* of contests wherein the first selects between V_1 and V_2 and each subsequent ballot pits the winner of the previous contest against the next proposal.¹²

The exhaustive approach requires $h \cdot (h-1)/2$ ballots B i.e. compare the first proposal to the $h-1$ others, compare the second to the $h-2$ remaining, $\dots = (h-1)+(h-2)+\dots+1 = h \cdot (h-1)/2$. This process may be much too exhausting when the number of alternatives is greater than 5 or 6 B e.g. $h=5$ requires 10 ballots, $h=6$ requires 15 ballots, $h=10$ requires 45 ballots, $h=20$ requires 190 ballots. Still, the exhaustive approach is sometimes feasible, so let us consider it first.

To apply either of the previous Jury Theorems (1 or 2) to a pairwise contest among V_i and V_j , let 'B' in the theorem be the claim that one of them, say V_i , is better than the other, V_j . The Jury Theorem then specifies a lower bound, p , on the group's competence level for each such pairwise contest. If the group competence level differs from contest to contest, let p be the smallest such lower bound. The best proposal of the lot, call it V^* , is quite likely to beat all of its rivals. How likely? If the group competence level is (at least) p for each pairwise contest, then under a fairly weak assumption the probability that the best proposal will defeat every alternative is at least p^{h-1} . The assumption we need is this. For each alternative V_j to the best proposal V^* :

The probability that V^* will defeat V_j is both: (1) probabilistically independent of whether V^* would defeat V_j 's rivals on other ballots, given that V^* is in fact better than every rival; and (2) probabilistically independent of the fact that V^* is better than V_j 's other rivals, given that V^* is better than V_j .

This clause says that the likelihood that V^* will defeat V_j depends on the fact that V^* is better than V_j , not

on how V^* may fare on votes against other proposals, and not on whether V^* is better than other proposals.

We need just a bit more notation to state the extensions of the Jury Theorems precisely. Let ' $\%v_j > 1/2$ ' say that in the contest between proposal V^* and alternative V_j a majority of votes favors V^* ; and let ' $V^* > V_j$ ' represent the claim that V^* is a better option than V_j . Then, the probability that the best proposal defeats each rival is this: $P[\%v_1 > 1/2 \cdots \%v_h > 1/2 \mid V^* > V_1 \cdots V^* > V_h] = P[\%v_1 > 1/2 \mid V^* > V_1 \cdots V^* > V_h] \cdots P[\%v_h > 1/2 \mid V^* > V_1 \cdots V^* > V_h] = P[\%v_1 > 1/2 \mid V^* > V_1] \cdots P[\%v_h > 1/2 \mid V^* > V_h] \geq p^{h-1}$, where each probability, $P[\%v_j > 1/2 \mid V^* > V_j] \geq p$, comes directly from the Jury Theorem (1 or 2) for binary decisions.

Consider a few examples. Suppose that in each contest between V^* and its alternatives, $r = .55$, $\text{cov} \leq s^2/(n-1)$, and each $\%v_j$ is Normally distributed (so Jury Theorem 2 applies). Table 2 says that for $n = 536$ voters we get $p = .99$ for each contest. With p at this level, if $h = 5$ proposals, the probability that the best proposal will defeat the other four is $(.99)^4 \approx .96$; and if $h = 10$, $p^{h-1} = (.99)^9 \approx .91$. If the circumstances are the same except that the population consists of $n = 945$ voters, then $p = .999$ (from Table 2). With p at this level, if $h = 5$ proposals, the probability that the best option will defeat all others is $(.999)^4 \approx .996$; and if $h = 10$, p^{h-1} will be $(.999)^9 \approx .991$. But also notice that if the group competence levels are only $p = .90$ on the pairwise comparisons, the likelihood that the best proposal will defeat four alternatives falls to $(.90)^4 = .66$; and for $h = 10$ it falls to $(.90)^9 \approx .39$. So, to maintain a high level of group competence at selecting the best among multiple alternatives the group competence level for pairwise comparisons must be quite high.

As a consequence of all this, a high group competence level p for pairwise comparisons makes it very unlikely that *cyclic majorities* will arise. If voters vote their *individual interests* it may quite plausibly happen that a majority will prefer V_1 to V_2 , another majority will prefer V_2 to V_3 , and yet a third majority will prefer V_3 to V_1 . But in contexts where a large enough population of moderately competent voters are in pursuit of the policy that will best promote the public good, the Jury Theorems imply that the group competence level, p , may be great enough that the best proposal will very probably defeat all others in pairwise contests, since p^{h-1} will remain quite high. Then the probability that a cyclic majority will arise is no greater than $1-p^{h-1}$, since a cycle can only occur if the best proposal loses to at least one alternative.

Even if it is feasible to run $h \cdot (h-1)/2$ separate ballots, in contexts where the best option is very likely to defeat each alternative there is little point in conducting such *exhaustive* pairwise comparisons. A *one-pass* series of contests will prove quite sufficient. The usual *legislative amendment process* is just such a *one-pass* series: a proposal, V_1 , is offered; an amendment, V_2 , is offered in its place (V_2 may be so different from V_1 as to hardly merit the term ‘amendment’); a vote determines whether V_2 defeats V_1 , and the winner may then face a new amendment V_3 . The process continues until the surviving proposal is submitted to a vote against the last alternative V_h (which is typically the proposal that the *status quo* be maintained, and no new policy be adopted). The total number of separate ballots is just $(h-1)$. If the group competence level is at least p for each pairwise comparison, then the probability that the best proposal will be adopted is *at least* p^{h-1} . The “worst case” occurs when the best proposal is among the first pair on offer, V_1 or V_2 , and must then endure pairwise contests against all $(h-1)$ alternative proposals. If, however, the best proposal is the last on offer, V_h , then the probability that it will be adopted is just p . If on average the best option first comes to a vote about half way through the process, then the group competence level at selecting it from among h proposals will be about $p^{(h-1)/2}$.

6.2 The Convincing Majorities Theorem for Multiple Alternatives. Theorem 3 described how the vote count should transform a person’s degree of confidence, and may convince her that the winner is indeed the better alternative. The present subsection shows how Theorem 3 may be extended to a *one-pass series* of contests among multiple alternatives, with much the same effect as in the simple binary case. However, the extension of Theorem 3 will not be quite so simple as the extension of the Jury Theorems just given. The notational conventions required to express the extended version of Theorem 3 makes this subsection rather technical. The reader who is not technically inclined may skip to Section 7, the concluding section of the paper.

Theorem 3 presupposes that the agent α has estimates of a lower bound on the average individual competence level, r , and an upper bound on the measure of how independently people vote, cov . To extend Theorem 3 to a *one-pass series* of contests we employ these same presuppositions and a bit more. But first we must introduce some additional notation.

Let V_1, \dots, V_h be the series of options in the order that they come to a vote. For the contest between V_1 and V_2 , call the winner ‘ W_1 ’ and the loser ‘ L_1 ’; let $\%w_1 = m_1/n$ be the fraction of votes for W_1 ; and $k_1 = m_1 - (n - m_1)$ is the number of votes by which W_1 wins. Whichever option, V_1 or V_2 , becomes the winner W_1 , it then challenges V_3 on the second ballot. The winner of that contest is called ‘ W_2 ’ and the loser is ‘ L_2 ’, and so on. For each of the $h-1$ ballots, W_i is the winner of the i^{th} ballot and L_i the loser; $\%w_i = m_i/n$ is the fraction of votes for W_i ; $k_i = m_i - (n - m_i)$ is the number of votes by which W_i wins. The proposal that wins the last round, W_{h-1} , is adopted. Let us also call this ultimate winner ‘ V^+ ’.

Now consider the losing proposals L_i ($1 \leq i \leq h-1$). There are $h-1$ of them, so there are $(h-1)!$ possible orderings among the losers. Let each such ordering represent a complete hypothesis about the ability of these alternatives to promote the public good, where lower ranked members are assessed by the ordering as inferior to higher ranked members. It will be useful to assign a number to each such ordering hypothesis; this number will merely be a label, with no other significance. So suppose each ordering is assigned a unique number from 1 through $(h-1)!$ and let ‘ O_s ’ represent ordering s . To turn a given ordering O_s into an ordering over *all* proposals one need only insert V^+ into it at some location j . So, for each O_s let ‘ O_s^j ’ denote the ordering of all proposals that results from placing V^+ at the j^{th} location of O_s and pushing the j^{th} and all “better alternatives” (according to O_s) up one position (i.e. into the $(j+1)^{\text{st}}$ and higher positions).

An ordering O_s^j represents a hypothesis about the relative effectiveness of each pair of alternatives at promoting the public good. We assume (for simplicity) that each proposal is either better or worse than its alternatives. So an ideally rational agent α should assign probabilities to these orderings that sum to 1:

$\sum_{s=1}^{(h-1)!} P_\alpha[O_s] = 1$, and $\sum_{s=1}^{(h-1)!} \sum_{j=1}^h P_\alpha[O_s^j] = 1$. The hypothesis that the ultimate winner V^+ is the best alternative is ‘ $V^+ > L_1 \cdot \dots \cdot V^+ > L_{h-1}$ ’, and it is equivalent to the disjunction of all of the hypotheses O_s^h that rate V^+ as *best*. Thus, $P_\alpha[V^+ > L_1 \cdot \dots \cdot V^+ > L_{h-1}] = \sum_{s=1}^{(h-1)!} P_\alpha[O_s^h]$.

The evidence that V^+ is best consists only of the vote counts from the binary contests, the data from each contest i that winner W_i received m_i votes and loser L_i received $n - m_i$ votes. In extending Theorem 3 it will turn out to be especially useful to track the hypotheses O_s that say that the winner W_i of the i^{th} ballot is in fact a *poorer* alternative than the loser L_i . For each O_s , let $\#(O_s)$ be the set of ballots i in the series such

that O_s rates the winner W_i as *worse* than the loser L_i . For $k_i = m_i - (n - m_i)$, the number of votes by which W_i defeats L_i on ballot i , the sum, $\sum_{i \in \#(O_s)} k_i$, is the total number of votes by which those proposals that O_s rates as worse than their direct rivals defeat those rivals in pairwise contests.

Similarly, for each complete hypothesis O_s^j , let $\#(O_s^j)$ be the set of ballots i in which O_s^j rates the ultimate winner V^+ as a *worse* alternative than its direct rival L_i on ballot i . The sum $\sum_{i \in \#(O_s^j)} k_i$ is the total number of votes by which the overall winner V^+ turns out to defeat those direct rivals rated by O_s^j to be better than V^+ . Both $\sum_{i \in \#(O_s^j)} k_i$ and $\sum_{i \in \#(O_s)} k_i$ will be central to the extension of Theorem 3.

Now we are almost ready to state the extension of Theorem 3 to multiple proposals. As in the extension of Theorems 1 and 2, we employ two straightforward independence assumptions. They are:

- (1) Given a complete hypothesis O_s^j about the relative merit of every pair of options, the probability that V_i will defeat V_k by some specific number of votes is *probabilistically independent* of the numbers of votes by which other pairs of proposals may defeat one another in other binary contests.
- (2) Given the hypothesis that V_i is better than V_k , the probability that V_i will defeat V_k by a specific number of votes is *probabilistically independent* of hypotheses about the relative merit of *other* pairs of options.

These assumptions say that for each pair of direct rivals, the likelihood that one of them, V_i , will defeat the other, V_k , by a specific number of votes probabilistically depends only on *their* relative merits, not on how votes on other proposals may turn out, and not on how the relative merits of other proposals stack up.

Let the expression ' $O_s^j(W_i, L_i)$ ' stand for ' $W_i > L_i$ ' or ' $L_i > W_i$ ', depending on whether the ordering O_s^j says that W_i is better than L_i or worse than L_i . Then the previous two independence assumptions imply that $P[\%w_1 = m_1/n \cdot \dots \cdot \%w_{h-1} = m_{h-1}/n \mid O_s^j] = P[\%w_1 = m_1/n \mid O_s^j] \cdot \dots \cdot P[\%w_{h-1} = m_{h-1}/n \mid O_s^j] = P[\%w_1 = m_1/n \mid O_s^j(W_1, L_1)] \cdot \dots \cdot P[\%w_{h-1} = m_{h-1}/n \mid O_s^j(W_{h-1}, L_{h-1})]$. From this equality the following extension of Theorem 3 follows easily. It employs a version of Bayes' Theorem that says, for evidence E and alternative hypotheses H_j , $P_a[H_1 \mid E] = 1/(1 + R)$, where $R = \sum_{j \neq 1} P[E \mid H_j] \cdot P_a[H_j] / P[E \mid H_1] \cdot P_a[H_1]$.

Theorem 3*. Extension of *Convincing Majorities* Theorem to Multiple Proposals. (Proved in Appendix.)

Suppose the assumptions of Theorem 3 hold for each ballot in a *one-pass* series of ballots on proposals V_1, \dots, V_h , and that the independence conditions just stated hold. Then the probability for person α that the winning proposal, V^+ , is the best alternative, given the vote count on each ballot is this:

$P_\alpha[V^+ > L_1 \cdot \dots \cdot V^+ > L_{h-1} \mid \%_{W_1} = m_1/n \cdot \dots \cdot \%_{W_{h-1}} = m_{h-1}/n] = 1/(1 + R)$, where

$$R = \frac{\sum_{s=1}^{(h-1)!} \exp[-(\sum_{i \in \#(O_s)} k_i) \cdot (r-1/2)/(n \cdot \sigma^2)] \cdot \sum_{j=1}^{h-1} \exp[-(\sum_{i \in \#(O_s^j)} k_i) \cdot (r-1/2)/(n \cdot \sigma^2)] \cdot P_\alpha[O_s^j]}{\sum_{s=1}^{(h-1)!} \exp[-(\sum_{i \in \#(O_s)} k_i) \cdot (r-1/2)/(n \cdot \sigma^2)] \cdot P_\alpha[O_s^h]}$$

Notice, if $\text{cov} \approx 0$, then $n \cdot \sigma^2 \approx r \cdot (1-r) - s^2$; and if cov is at least small enough that $(n-1) \cdot \text{cov} \leq s^2$, then $n \cdot \sigma^2 \leq r \cdot (1-r)$. In either case the number, n , of voters is irrelevant. Only the numbers of votes by which winners defeat higher rated losers (i.e. $\sum_{i \in \#(O_s)} k_i$ and $\sum_{i \in \#(O_s^j)} k_i$), together with the value of the average competence level r (and perhaps the value of s^2) play a significant role.

As R goes to 0, the posterior probability (for agent α) that the ultimate winner V^+ is the best proposal goes to 1. So we will briefly investigate the circumstances under which R may become small. For a specific sequence of outcomes from ballots one may compute the value of R directly. But it will be useful to get a handle on how R tends to behave in general.

Notice that in the formula for R each ordering O_s that rates all winners W_i as better than losers L_i will make the set $\#(O_s)$ empty; so $\sum_{i \in \#(O_s)} k_i = 0$. And an ordering O_s^j that rates V^+ better than each of its direct rivals makes $\#(O_s^j)$ empty, so $\sum_{i \in \#(O_s^j)} k_i = 0$. In both cases the whole term inside the corresponding exp function is 0, and $\exp[0] = 1$. But when O_s rates at least one winner as worse than a loser, the term $\sum_{i \in \#(O_s)} k_i$ will be greater than 1 and the whole term inside exp is negative; and when x is negative, $0 < \exp[x] < 1$. A similar observation applies to the terms involving $\#(O_s^j)$.

To get a clearer picture of how R works it may help to consider a few special cases. For $h = 2$ there are only two proposals, V_1 and V_2 . One will be the winner V^+ and the other the loser L_1 . There is just one ordering O_1 that does not contain the winner V^+ . It contains only L_1 . O_1^1 has V^+ rated below L_1 , and O_1^2 has V^+ rated above L_1 . In this case it is easily verified that the equation for the posterior probability reduces to that given in Theorem 3.

Suppose there are more than two proposals, but the ultimate winner is the first (i.e. V^+ is V_1). So V^+ must contend with each alternative. For each hypothesis O_s^j with $j < h$, the term $\sum_{i \in \#(O_s^j)} k_i$ is positive. If each k_i is a *convincing majority* (as in Theorem 3) on ballot i , then all factors of form $\exp[-(\sum_{i \in \#(O_s^j)} k_i) \cdot (r-1/2)/(n \cdot \sigma^2)]$ must be extremely small fractions; they will be extremely small when $j = h-1$, and will be exponentially smaller for larger values of j . So, even if prior probabilities of hypotheses that say V^+ is not the best option are much larger than priors for hypotheses that say V^+ is the best (i.e. even if for $j \geq 2$, $P_a[O_s^j] \gg P_a[O_s^1]$), the factor R will be made extremely small by the convincing majorities k_i that result from the binary contests.

More generally, for a hypothesis O_s that says some proposals are superior to rivals that defeated them, the set $\#(O_s)$ is non-empty and the term $\exp[-(\sum_{i \in \#(O_s)} k_i) \cdot (r-1/2)/(n \cdot \sigma^2)]$ may be extremely small. However, for an O_s that orders proposals in a way consistent with their win-loss tally, the set $\#(O_s)$ is empty and the term $\exp[-(\sum_{i \in \#(O_s)} k_i) \cdot (r-1/2)/(n \cdot \sigma^2)]$ equals 1. Thus, hypotheses O_s that contribute most to the positive sizes of the numerator and denominator of R are those that rate winners W_i higher than losers L_i . Call these the *main contributors*. The extension of a *main contributor* O_s to a hypothesis O_s^j for $j < h$ must rate V^+ below its last rival; and when $j < h-1$, O_s^j will rate V^+ below several rivals. Thus, for any *main contributor* O_s , the factor $\exp[-(\sum_{i \in \#(O_s^j)} k_i) \cdot (r-1/2)/(n \cdot \sigma^2)]$ that multiplies the prior probability $P_a[O_s^j]$ in the numerator of R is extremely small, but the factor that multiplies $P_a[O_s^h]$ in the denominator of R is 1. All of this contributes to making R small. As a result the posterior probability that V^+ is the best alternative becomes quite large.

Thus, in the evaluation of proposals through the legislative amendment process, if the winner of each binary contest defeats the loser by a *convincing majority*, then the posterior probability that the ultimate winner is the best among them will be close to 1. And as in the simple binary case, a high posterior probability for the winner depends only on the absolute differences between numbers of votes pro and con.

7. Conclusion

Whenever the ballot is employed to select policies or laws or office holders in order to promote the public good, the Jury Theorems apply. These theorems demonstrate just how reliable democratic voting can be. They also reveal the extent to which group competence suffers as average voter competence

decreases. And they show precisely how voting in blocks diminishes group competence by, in effect, decreasing the size of the electorate. Thus, Jury Theorems model aspects of voting that are critical to the epistemology of majority judgments.

Modern democracies may be ill suited to secure the benefits that accrue to groups of moderately competent, independently-minded voters. The average individual competence levels of voters and legislators at discerning the better policy may fall short. They may too often vote their private interests or the interests of influential friends and constituents. And modern democracies may fall too much under the sway of parties and factions to secure the high group competence levels that would attend a similarly sized body of more independently-minded voters. In this regard Jury Theorem models are like other formal models of complex social institutions. They provide insights into the roles institutions may play, and reveal their capacities and limitations. And they may suggest ways to improve the institutions they model.

Appendix

We begin with a brief list of formal Definitions used in the paper:

For all i and $j \neq i$: $r_i = P[b_i=1 \mid B]$, $1-r_i = P[b_i=0 \mid B]$, $r_{ij} = P[b_i=1 \cdot b_j=1 \mid B]$.

For $n =$ the number of voters: $r = \sum_{i=1}^n r_i/n$, $s^2 = \sum_{i=1}^n (r_i-r)^2/n$, $\text{cov} = [2/(n \cdot (n-1))] \cdot \sum_{i=1}^{n-1} \sum_{j=i+1}^n (r_{ji}-r_j \cdot r_i)$,

$$r^* = \sum_{k=0}^n (k/n) \cdot P[\%b = k/n \mid B], \quad \sigma^2 = \sum_{k=0}^n ((k/n) - r^*)^2 \cdot P[\%b = k/n \mid B].$$

The definitions of r^* and σ^2 imply the following (by grouping, for each k , the ways in which $(\sum_{i=1}^n v_i) = k$):

$$r^* = \sum_{v_1=0}^1 \dots \sum_{v_n=0}^1 ((\sum_{i=1}^n v_i)/n) \cdot P[b_1=v_1 \cdot \dots \cdot b_n=v_n \mid B]$$

$$\sigma^2 = \sum_{v_1=0}^1 \dots \sum_{v_n=0}^1 ((\sum_{i=1}^n v_i)/n - r^*)^2 \cdot P[b_1=v_1 \cdot \dots \cdot b_n=v_n \mid B].$$

Proof that $r^* = r$:

$$r^* = (1/n) \cdot \sum_{v_1=0}^1 \dots \sum_{v_n=0}^1 (\sum_{i=1}^n v_i \cdot P[b_1=v_1 \cdot \dots \cdot b_n=v_n \mid B])$$

$$= (1/n) \cdot \sum_{v_1=0}^1 \dots \sum_{v_{n-1}=0}^1 \{ (\sum_{i=1}^{n-1} v_i \cdot P[b_1=v_1 \cdot \dots \cdot b_{n-1}=v_{n-1} \cdot b_n=0 \mid B]) +$$

$$(\sum_{i=1}^{n-1} v_i \cdot P[b_1=v_1 \cdot \dots \cdot b_{n-1}=v_{n-1} \cdot b_n=1 \mid B]) + P[b_1=v_1 \cdot \dots \cdot b_{n-1}=v_{n-1} \cdot b_n=1 \mid B] \}$$

$$= (1/n) \cdot \sum_{v_1=0}^1 \dots \sum_{v_{n-1}=0}^1 (\sum_{i=1}^{n-1} v_i \cdot P[b_1=v_1 \cdot \dots \cdot b_{n-1}=v_{n-1} \mid B]) +$$

$$(1/n) \cdot \sum_{v_1=0}^1 \dots \sum_{v_{n-1}=0}^1 P[b_1=v_1 \cdot \dots \cdot b_{n-1}=v_{n-1} \cdot b_n=1 \mid B]$$

$$\begin{aligned}
&= (1/n) \cdot \sum_{v_1=0}^1 \dots \sum_{v_{n-1}=0}^1 (\sum_{i=1}^{n-1} v_i \cdot P[b_1=v_1 \cdot \dots \cdot b_{n-1}=v_{n-1} \mid B]) + (1/n) \cdot P[b_n=1 \mid B] \\
&= \dots = (1/n) \cdot P[b_1=1 \mid B] + \dots + (1/n) \cdot P[b_n=1 \mid B] = (1/n) \cdot r_1 + \dots + (1/n) \cdot r_n = r
\end{aligned}$$

Proof that $\sigma^2 = [r(1-r)/n] - [s^2/n] + ((n-1)/n) \cdot \text{cov}$:

The proof will employ the equivalence derived in the following lines:

$$\begin{aligned}
&\sum_{v_i=0}^1 \sum_{v_j=0}^1 (v_i - r_i) \cdot (v_j - r_j) \cdot P[b_i=v_i \cdot b_j=v_j \mid B] \\
&= (0-r_i) \cdot (0-r_j) \cdot P[b_i=0 \cdot b_j=0 \mid B] + (0-r_i) \cdot (1-r_j) \cdot P[b_i=0 \cdot b_j=1 \mid B] + \\
&\quad (1-r_i) \cdot (0-r_j) \cdot P[b_i=1 \cdot b_j=0 \mid B] + (1-r_i) \cdot (1-r_j) \cdot P[b_i=1 \cdot b_j=1 \mid B] \\
&= (P[b_i=0 \cdot b_j=0 \mid B] + P[b_i=0 \cdot b_j=1 \mid B] + P[b_i=1 \cdot b_j=0 \mid B] + P[b_i=1 \cdot b_j=1 \mid B]) \cdot r_i \cdot r_j - \\
&\quad (P[b_i=0 \cdot b_j=1 \mid B] + P[b_i=1 \cdot b_j=0 \mid B]) \cdot r_i - (P[b_i=1 \cdot b_j=0 \mid B] + P[b_i=1 \cdot b_j=1 \mid B]) \cdot r_j + P[b_i=1 \cdot b_j=1 \mid B] \\
&= r_i \cdot r_j - r_j \cdot r_i - r_i \cdot r_j + r_{ij} = r_{ij} - r_i \cdot r_j. \text{ Thus, } \sum_{v_i=0}^1 \dots \sum_{v_j=0}^1 (v_i - r_i) \cdot (v_j - r_j) \cdot P[b_i=v_i \cdot b_j=v_j \mid B] = (r_{j,i} - r_j \cdot r_i). \\
\sigma^2 &= (1/n)^2 \cdot \sum_{v_i=0}^1 \dots \sum_{v_n=0}^1 (\sum_{i=1}^n (v_i - r_i))^2 \cdot P[b_1=v_1 \cdot \dots \cdot b_n=v_n \mid B] \\
&= (1/n)^2 \cdot \sum_{v_i=0}^1 \dots \sum_{v_n=0}^1 [\sum_{i=1}^n (v_i - r_i)^2 + 2 \cdot \sum_{i=1}^{n-1} \sum_{j=i+1}^n (v_i - r_i) \cdot (v_j - r_j)] \cdot P[b_1=v_1 \cdot \dots \cdot b_n=v_n \mid B] \\
&= (1/n)^2 \cdot \sum_{i=1}^n \{ \sum_{v_i=0}^1 (v_i - r_i)^2 \cdot (\sum_{v_1=0}^1 \dots \sum_{v_{i-1}=0}^1 \sum_{v_{i+1}=0}^1 \dots \sum_{v_n=0}^1 P[b_1=v_1 \cdot \dots \cdot b_n=v_n \mid B]) \} + \\
&\quad 2 \cdot (1/n)^2 \cdot \sum_{i=1}^{n-1} \sum_{j=i+1}^n \{ \sum_{v_i=0}^1 \sum_{v_j=0}^1 (v_i - r_i) \cdot (v_j - r_j) \cdot \\
&\quad (\sum_{v_1=0}^1 \dots \sum_{v_{i-1}=0}^1 \sum_{v_{i+1}=0}^1 \dots \sum_{v_{j-1}=0}^1 \sum_{v_{j+1}=0}^1 \dots \sum_{v_n=0}^1 P[b_1=v_1 \cdot \dots \cdot b_n=v_n \mid B]) \} \\
&= (1/n)^2 \cdot \sum_{i=1}^n \sum_{v_i=0}^1 (v_i - r_i)^2 \cdot P[b_i=v_i \mid B] + 2 \cdot (1/n)^2 \cdot \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{v_i=0}^1 \sum_{v_j=0}^1 (v_i - r_i) \cdot (v_j - r_j) \cdot P[b_i=v_i \cdot b_j=v_j \mid B] \\
&= (1/n)^2 \cdot \sum_{i=1}^n r_i \cdot (1-r_i) + 2 \cdot (1/n)^2 \cdot \sum_{i=1}^{n-1} \sum_{j=i+1}^n (r_{ij} - r_i \cdot r_j) \\
&= (1/n) \cdot [r - (1/n) \cdot \sum_{i=1}^n r_i^2] + [(n-1)/n] \cdot 2 / (n \cdot (n-1)) \cdot \sum_{i=1}^{n-1} \sum_{j=i+1}^n (r_{j,i} - r_j \cdot r_i) \\
&= (1/n) \cdot [r(1-r) + r^2 - (1/n) \cdot \sum_{i=1}^n r_i^2] + ((n-1)/n) \cdot \text{cov} \\
&= [r(1-r)/n] - (1/n)^2 \cdot \sum_{i=1}^n (r_i^2 - r^2) + ((n-1)/n) \cdot \text{cov} \\
&= [r(1-r)/n] - (1/n)^2 \cdot \sum_{i=1}^n (r_i - r)^2 + ((n-1)/n) \cdot \text{cov} \\
&= [r(1-r)/n] - [s^2/n] + ((n-1)/n) \cdot \text{cov}
\end{aligned}$$

Proof of **Theorem 1**, the *Weak Law* Jury Theorem:

Clause (1) of Theorem 1 may be derived as follows, for each small $\varepsilon > 0$:

$$\begin{aligned}
\sigma^2 &= \sum_{k=0}^n ((k/n) - r)^2 \cdot P[\#b = k/n \mid B] \\
&= (1/n)^2 \cdot \sum_{k=0}^{\lfloor (r-\varepsilon)n \rfloor + 1} (k-n \cdot r)^2 \cdot P[\#b=k \mid B] + (1/n)^2 \cdot \sum_{k=\lfloor (r-\varepsilon)n \rfloor + 2}^{\lfloor (r+\varepsilon)n \rfloor - 1} (k-n \cdot r)^2 \cdot P[\#b=k \mid B] +
\end{aligned}$$

$$\begin{aligned}
& (1/n)^2 \cdot \sum_{k=[(r+\varepsilon)\cdot n]}^n (k-n\cdot r)^2 \cdot P[\#b=k \mid B] \\
& \geq (1/n)^2 \cdot (\varepsilon\cdot n)^2 \cdot (\sum_{k=0}^{[(r-\varepsilon)\cdot n]} P[\#b=k \mid B] + \sum_{k=[(r+\varepsilon)\cdot n]}^n P[\#b=k \mid B]) \\
& \geq \varepsilon^2 \cdot (P[0 \leq \#b \leq (r-\varepsilon)\cdot n \mid B] + P[(r+\varepsilon)\cdot n \leq \#b \leq n \mid B]).
\end{aligned}$$

$$\text{So, } \sigma^2/\varepsilon^2 \geq 1 - P[(r-\varepsilon) < \%b < (r+\varepsilon) \mid B] = 1 - P[-\varepsilon < \%b < \varepsilon \mid B]$$

Clause (2.1) may be derived as follows (and clause (2.2) is proved similarly):

Suppose $r > 1/2$ and let $\alpha = \sigma^2/(r - 1/2)$. Then, $[\sigma^2/(r - 1/2)^2] \cdot [(r - 1/2)^2 + \sigma^2] = \sigma^2 + \alpha^2$, and

$$\begin{aligned}
\sigma^2 + \alpha^2 &= \sigma^2 - 2\cdot\alpha\cdot(r-r) + \alpha^2 \\
&= \sum_{k=0}^n ((k/n) - r)^2 \cdot P[\#b=k \mid B] - \sum_{k=0}^n 2\cdot\alpha\cdot((k/n) - r) \cdot P[\#b=k \mid B] + \sum_{k=0}^n \alpha^2 \cdot P[\#b=k \mid B] \\
&= \sum_{k=0}^n [((k/n) - r)^2 - 2\cdot\alpha\cdot((k/n) - r) + \alpha^2] \cdot P[\#b=k \mid B] \\
&= \sum_{k=0}^n ((k/n) - r - \alpha)^2 \cdot P[\#b=k \mid B] \\
&\geq \sum_{k=0}^{n/2} ((k/n) - r - \alpha)^2 \cdot P[\#b=k \mid B] \\
&\geq (1/2 - r - \alpha)^2 \cdot \sum_{k=0}^{n/2} P[\#b=k \mid B] \\
&= ((r - 1/2)^2 + \alpha^2 + 2\cdot\alpha\cdot(r - 1/2)) \cdot P[\%b \leq 1/2 \mid B] \\
&= [((r - 1/2)^2 + \sigma^2)^2 / (r - 1/2)^2] \cdot P[\%b \leq 1/2 \mid B].
\end{aligned}$$

$$\text{So, } \sigma^2 \geq ((r - 1/2)^2 + \sigma^2) \cdot (1 - P[\%b > 1/2 \mid B]).$$

$$\text{Thus, } P[\%b > 1/2 \mid B] \geq 1 - [\sigma^2/((r - 1/2)^2 + \sigma^2)] = 1 / (1 + \sigma^2/(r-1/2)^2).$$

Proof of Theorem 3, the Convincing Majorities Theorem:

Clause (1) follows from the fact that if the distribution of $\%b$ (given B) is nearly Normal,

$$P[\%b = m/n \mid B] = (2\pi\cdot\sigma^2)^{-1/2} \cdot \int_{(m-1/2)/n}^{(m+1/2)/n} \exp[-(x-r)^2/2\sigma^2] dx \approx (2\pi\cdot\sigma^2)^{-1/2} \cdot (1/n) \cdot \exp[-(m/n - r)^2/2\sigma^2].$$

Clauses (2) and (2') give *likelihood ratios* that result from clause (1), i.e., the likelihood that $\%b = m/n$

(so $\%c = (n-m)/n$) if C holds, divided by the likelihood that $\%b = m/n$ if B holds. The derivation is this:

$$\begin{aligned}
P[\%c=(n-m)/n \mid C] / P[\%b=m/n \mid B] &\approx \exp[-((n-m)/n - r)^2/2\sigma^2] / \exp[-((m/n) - r)^2/2\sigma^2] \\
&= \exp[[(m - nr)^2 - ((n-m) - nr)^2] / (2\cdot n^2 \cdot \sigma^2)] \\
&= \exp[(m^2 - 2mnr + (nr)^2 - n^2 + 2mn - m^2 + 2n^2r - 2mnr - (nr)^2) / (2\cdot n^2 \cdot \sigma^2)] \\
&= \exp[(-4mnr + n^2 + 2mn + 2n^2r) / (2\cdot n^2 \cdot \sigma^2)] \\
&= \exp[-(2\cdot m - n) \cdot (r - 1/2) / (n \cdot \sigma^2)]
\end{aligned}$$

$$= \exp[-k \cdot (r - 1/2) / (n \cdot \sigma^2)]$$

$$\leq \exp[-k \cdot (r - 1/2) / r \cdot (1-r)] \text{ when } \text{cov} \leq s^2/(n-1).$$

Likelihood ratios play a crucial role in the relationship between prior and posterior probabilities given by Bayes' Theorem. Clauses (3) and (3') are forms of Bayes' Theorem, derivable from the following:

$$P_a[B \mid \%b = m/n] = P[\%b = m/n \mid B] \cdot P_a[B] / (P[\%b = m/n \mid B] \cdot P_a[B] + P[\%b = m/n \mid C] \cdot P_a[C]).$$

Proof of **Theorem 3***, the Extension of the *Convincing Majorities* Theorem to Multiple Proposals:

First, notice that for each i ($1 \leq i \leq h-1$), for $k_i = m_i - (n - m_i)$, it follows from the proof of Theorem 3 that:

$$\text{if } O_s^j(W_i, L_i) \text{ says ' } W_i > L_i \text{ ', then } P[\%w_i = m_i/n \mid O_s^j(W_i, L_i)] = \exp[-(m_i/n - r)^2/2\sigma^2] / (2\pi)^{1/2} \cdot \sigma \cdot n ;$$

$$\text{if } O_s^j(W_i, L_i) \text{ says ' } L_i > W_i \text{ ', then } P[\%w_i = m_i/n \mid O_s^j(W_i, L_i)] = \exp[-((n - m_i)/n - r)^2/2\sigma^2] / (2\pi)^{1/2}$$

$$= \exp[-k_i \cdot (r - 1/2) / (n \cdot \sigma^2)] \cdot \exp[-(m_i/n - r)^2/2\sigma^2] / (2\pi)^{1/2} \cdot \sigma \cdot n .$$

$$\text{Thus, for each } s \text{ and } j, P[\%w_1 = m_1/n \cdot \dots \cdot \%w_{h-1} = m_{h-1}/n \mid O_s^j] = \prod_{i=1}^{h-1} P[\%w_i = m_i/n \mid O_s^j(W_i, L_i)]$$

$$= \prod_{i=1}^{h-1} \{ \exp[-(m_i/n - r)^2/2\sigma^2] / (2\pi)^{1/2} \cdot \sigma \cdot n \} \cdot \prod_{i \in \#(O_s)} \exp[-k_i \cdot (r - 1/2) / (n \cdot \sigma^2)] \cdot$$

$$\prod_{i \in \#(O_s)} \exp[-k_i \cdot (r - 1/2) / (n \cdot \sigma^2)]$$

$$= \prod_{i=1}^{h-1} \{ \exp[-(m_i/n - r)^2/2\sigma^2] / (2\pi)^{1/2} \cdot \sigma \cdot n \} \cdot \exp[-(\sum_{i \in \#(O_s)} k_i) \cdot (r - 1/2) / (n \cdot \sigma^2)] \cdot$$

$$\exp[-(\sum_{i \in \#(O_s)} k_i) \cdot (r - 1/2) / (n \cdot \sigma^2)].$$

Bayes' Theorem yields $P_a[V^+ > L_1 \cdot \dots \cdot V^+ > L_{h-1} \mid \%w_1 = m_1/n \cdot \dots \cdot \%w_{h-1} = m_{h-1}/n] = 1/(1+R)$, $R =$

$$\sum_{s=1}^{(h-1)!} \sum_{j=1}^{h-1} P[\%w_1 = m_1/n \cdot \dots \cdot \%w_{h-1} = m_{h-1}/n \mid O_s^j] \cdot P_a[O_s^j] / P[\%w_1 = m_1/n \cdot \dots \cdot \%w_{h-1} = m_{h-1}/n \mid O_s^h] \cdot P_a[O_s^h].$$

References

Aristotle. 1971a. *Nicomachean Ethics*. In *The Basic Works of Aristotle*, trans. W. D. Ross, ed. R.

McKeon. Random House.

----. 1971b. *Politics*. In *The Basic Works of Aristotle*, trans. B. Jowett, ed. R. McKeon. Random House.

Arrow, K. 1951. *Social Choice and Individual Values*. Yale U. Press. 2d ed., 1963.

----. 1977. "Current Developments in the Theory of Social Choice." *Social Research* 44:607-622.

Reprinted in B. Barry, R. Hardin, eds., *Rational Man and Irrational Society?*, Sage, 1982.

Barry, B. 1964. "The Public Interest." *Proceedings of the Aristotelian Society* 38:1-18.

- Billingsley, P. 1986. *Probability and Measure*. Wiley.
- Black, D. 1958. *The Theory of Committees and Elections*. Cambridge U. Press. Republished, 1971.
- Christiano, T. 1996. *The Rule of the Many: Fundamental Issues in Democratic Theory*. Westview Press.
- Cohen, J. 1986. "An Epistemic Conception of Democracy." *Ethics* 97:26-38.
- Coleman, J., Ferejohn, J. 1968. "Democracy and Social Choice." *Ethics* 97:6-25.
- Condorcet, N. C. de. 1785. *Essai sur l'Application de l'Analyse a la Probabilite des Decisions Rendues a la Pluralite des voix*. Paris.
- Estlund, D. 1993. "Making truth safe for Democracy." In *The Idea of Democracy*, ed. D. Copp, J. Hampton, J. E. Roemer. Cambridge U. Press.
- . 1994. "Opinion Leaders, Independence, and Condorcet's Jury Theorem." *Theory and Decision* 36:131-162.
- Estlund, D., Waldron, J., Grofman, B., Feld, S. 1989. "Democratic Theory and the Public Interest, Condorcet and Rousseau Revisited." *American Political Science Review* 83:1317-1340.
- Feller, W. 1968. *An Introduction to Probability Theory and Its Applications*, vols. 1. Wiley.
- Feller, W. 1971. *An Introduction to Probability Theory and Its Applications*, vols. 2. Wiley.
- Goodin, R. 2002. "The Paradox of Persisting Opposition." *Politics, Philosophy and Economics* 1: 109-146.
- Grofman, B., Feld, S. 1988. "Rousseau's General Will." *American Political Science Review* 82:567-576.
- Grofman, B., Owen, G., eds. 1986. *Information Pooling and Group Decision-Making*. JAI Press.
- Grofman, B., Owen, G., Feld, S. 1983. "Thirteen Theorems in Search of the Truth." *Theory and Decision* 15: 261-278.
- Howson, C., Urbach, P. 1993. *Scientific Reasoning: the Bayesian Approach*. 2nd ed. Open Court.
- Ladha, K. 1992. "The Condorcet Jury Theorem, Free Speech, and Correlated Votes." *American Journal of Political Science* 36: 617-634.
- . 1995. "Information Pooling Through Majority-Rule Voting: Condorcet's Jury Theorem and Correlated Voting." *Journal of Economic Behavior and Organization* 26: 353-372.

- List, C., Goodin, R. 2001. "Epistemic Democracy: Generalizing the Condorcet Jury Theorem." *Journal of Political Philosophy* 9: 277-306.
- Locke, J. 1690. *Second Treatise on Civil Government*. London. In *Social Contract*, ed. E. Barker. Oxford U. Press, 1960.
- Mill, J. S. 1843. *A System of Logic*. London. Revised 1866, 1872.
- . 1859. *On Liberty*. London.
- . 1861. *Considerations on Representative Government*. London.
- . 1863. *Utilitarianism*. London.
- Nussbaum, M. 1992. "Human Functioning and Social Justice: In Defense of Aristotelian Essentialism." *Political Theory* 20: 202-246.
- Owen, G., Grofman, B., Feld, S. 1989. "Proving a Distribution-Free Generalization of the Condorcet Jury Theorem." *Mathematical Social Sciences* 17:1-16.
- Rawls, J. 1971. *A Theory of Justice*. Harvard U. Press.
- Riker, W. 1982. *Liberalism Against Populism*. W. H. Freeman and Company.
- Rousseau, J. J. 1762. *du Contract Social*. Paris. In *Social Contract*, trans. G.Hopkins, ed. E. Baker.Oxford U. Press, 1960.
- Sen, A. K. 1970. *Collective Choice and Social Welfare*. Holden-Day.
- . 1982. *Choice, Welfare, and Measurement*. MIT Press.
- . 1986. "Social Choice Theory." In *Handbook of Mathematical Economics, Volume III.*, ed. K. J. Arrow and M. D. Intriligator. North-Holland.
- . 1993. "Capability and Well-Being." In *The Quality of Life*, ed. M. Nussbaum and A. K. Sen. Oxford U. Press, 1993.
- Trachtenberg, Z. 1993. *Making Citizens: Rousseau's Political Theory of Culture*. Routledge.

NOTES

1. Perhaps the chief value of democratic voting does not derive from such instrumental benefits. E.g., Christiano (1996) argues that democracy is intrinsically valuable because citizen's interests have equal intrinsic worth and political equality is the best way to accommodate this fact. Even so, the question regarding voting's instrumental value remains: is it an effective means of selecting policy options that promote the public good, or may it fairly aggregate preferences, or does it mainly function as a check on government power?
2. See, e.g., Arrow (1951, 1977), Black, (1958), and Sen, (1970, 1982). Sen (1986) provides a comprehensive overview of social choice theory and the various formal results regarding the existence of social choice functions.
3. Condorcet described a version of the Jury Theorem in his (1785). Mill knew of Condorcet's work but questioned its applicability to democratic voting (1843, Bk.3, ch.18, sec.3). The theorem received little additional attention until Black's study (1958, pp.159-165). Black finds the theorem inapplicable to democratic voting on the grounds that in that context the notion of a correct opinion "seems to be without definite meaning." But his book sparked renewed interest in Jury Theorems. Barry (1964) argues that *the public interest* is a meaningful notion and finds support in Rousseau's notion of *the general will*. He draws on the Jury Theorem in support of Rousseau and his paper initiated an enduring association between Jury Theorems and Rousseau's views. Substantial work on Jury Theorems and their implications may be found in the following articles: Grofman, et al. (1983), Coleman and Ferejohn (1986), Cohen (1986), Grofman and Feld (1988), Estlund, Waldron, Grofman and Feld (1989), Owen, et al. (1989), Estlund (1994), and Ladha (1992, 1995). Also see the articles in Grofman and Owen (1986). The present paper builds on this work but does not presuppose that the reader is familiar with it.
4. See, e.g., the discussion of independence by Grofman and Feld (1988). But also see the remarks of Waldron and the reply by Grofman and Feld in (Estlund, et. al. 1989) for a more considered view.

5. For contemporary versions of Jury Theorems see (Grofman, et al., 1983), (Owen, et al., 1989), and (Ladha, 1992, 1995). The Jury Theorems I will present are the most general (i.e. strongest) versions applicable to simple majority decisions. They are applications of well known mathematical results -- i.e. versions of Chebyshev's inequality and the Central Limit Theorem together with characteristics of the Normal Distribution.
6. Nussbaum (1992) and Sen (1993) articulate significant contemporary Aristotelian conceptions of the nature of the public good.
7. Trachtenberg (1993) offers an insightful assessment of the affinity of Jury Theorems with Rousseau's views. Although he finds Jury Theorems congenial to much of what Rousseau says, he argues persuasively that Rousseau's view about the need for a *civil religion* to enforce obedience to the current laws is strongly at odds with the kind of critical search for better laws suggested by the Jury Theorem model of voting.
8. Points (1) and (2) will be treated thoroughly in any good text on probability and statistics. I suggest Feller (1968, 1971). On point (3) see Billingsley (1986).
9. This theorem follows easily from the basic properties of the Normal distribution. See, e.g., (Feller, 1968).
10. For a good overview of probabilistic rationality and Bayesian inference see (Howson and Urbach, 1993).
11. The issue of whether the minority should be convinced by the votes of the majority, as the Bayesian model suggests, or whether the persistence of opposition after the votes are counted may be rationally justified deserves a much more thorough treatment. See the paper by Robert Goodin (2002) for a really excellent analysis of this issue.
12. Another obvious method is to hold just one ballot and employ a plurality voting rule. The simplest

plurality rule permits each voter a single vote and adopts the policy that receives the most votes. List and Goodin (2001) establish conditions under which a Jury Theorem holds for such plurality voting. Their theorem applies to probabilistically independent voters who share the same multinomial distribution of voting for the various alternatives and where the best option proposed is more likely to receive an agent's vote than any of the alternatives.