14.770: Introduction to Political Economy
Lecture 2: When Voting Works and When It Doesn't: A First Look

Daron Acemoglu

MIT

September 13, 2022.

Introduction

- The Median Voter Theorem (MVT) is powerful (though very special), in part because it also creates strong incentives for parties and political candidates.
- But, as we will see in the next lecture, it has only limited success in describing how politics works in democracies.
- What happens away from the MVT?
- I will talk about two aspects of this question in today's lecture.
  - When there is uncertainty about how voters will vote.
  - When there is uncertainty about what policies are beneficial.
- To start with, let us first recap the logic of Downsian policy convergence

A Simple Model of Indirect Democracy

- Two parties that can announce and *commit* to policies.
- Rent $Q > 0$ from coming to power and no ideological bias.
- Thus the maximization problem of the two parties are

$$\begin{aligned} \text{Party } A &: \quad \max_{p_A} \mathbb{P}(p_A, p_B) Q \\ \text{Party } B &: \quad \max_{p_B} \left(1 - \mathbb{P}(p_A, p_B)\right) Q \end{aligned}$$

- $\mathbb{P}(p_A, p_B)$ is the probability that party A comes to power when the two parties' platforms are p_A and p_B respectively.

Party Competition

- Let the bliss point of the median voter be p_m.
- When the median voter theorem applies, we have

$$\begin{aligned} \mathbb{P}(p_A, p_B &= p_m) = 0, \ \mathbb{P}(p_A = p_m, p_B) = 1, \text{ and} \\ \mathbb{P}(p_A &= p_m, p_B = p_m) \in [0,1]. \end{aligned}$$

A4. *Randomization:*

$$\mathbb{P}(p_A = p_m, p_B = p_m) = 1/2.$$

Why?

Downsian Policy Convergence Theorem

Theorem

(Downsian Policy Convergence Theorem) *Suppose that there are two parties that first announce a policy platform and commit to it and a set of voters* $\mathcal{H}$ *that vote for one of the two parties. Assume that A4 holds and that all voters have single-peaked policy preferences over a given ordering of policy alternatives, and denote the median-ranked bliss point by* p_m. *Then, both parties will choose* p_m *as their policy platform.*

Proof of the Downsian Policy Convergence Theorem

- The proof is by contradiction.
- Suppose not, then there is a profitable deviation for one of the parties.
- For example, if $p^A > p^B > p_m$, one of the parties can announce p_m and win the election for sure.
- When $p^A \neq p_m$ and $p^B = p_m$, party A can also announce p_m and increase its chance of winning to $1/2$.

Downsian Policy Convergence Theorem: Discussion

- What happens without Assumption A4?
- Why is this theorem important?
- A natural generalization of this theorem would be to consider three or more parties. What happens with three parties?

What Happens with Policy-Motivated Politicians?

- Instead office-seeking parties, suppose that parties/politicians can commit to policies, but have policy preferences.
- For example, one party may prefer right-wing policies the other one left-wing ones.
- What happens in this case?

Theorem

Suppose we are in the baseline model with single-peaked or single-crossing preferences, and the two parties have their own policy platforms, one to the left of the median the other one to the right of the median. The unique equilibrium is Dowsian policy convergence.

- Why?
- But this result as fragile as we will see.

Application: Redistributive Taxation I

- Consider situation with two parties competing to come to power.
- Suppose that agents have the following preferences

$$u^i\left(c^i, x^i\right) = c^i + h(x^i)$$

  where c^i and x^i denote individual consumption and leisure, and $h(\cdot)$ is a well-behaved concave utility function.
- There are only two policy instruments, linear tax on earnings τ on lump-sum transfers $T \geq 0$ (and this is important).
- The budget constraint of each agent is

$$c^i \leq (1 - \tau)l^i + T,$$

- The real wage is exogenous and normalized to 1.
- Individual productivity differs, such that the individuals have different amounts of "effective time" available. That is, individuals are subject to the "time constraint"

$$\alpha^i \geq x^i + l^i,$$

Application: Redistributive Taxation II

- Assume that α^i is distributed in the population with mean α and median α^m.
- Since individual preferences are linear in consumption, optimal labor supply satisfies
$$l^i = L(\tau) + (\alpha^i - \alpha),$$
where $L(\tau) \equiv \alpha - (h')^{-1}(1-\tau)$ is decreasing in τ by the concavity of $h(\cdot)$.
- To derive this, note that from quasi-linear preferences, the first order condition of each individual is
$$(1-\tau) = h'(x^i).$$
Inverting this, writing $x^i = \alpha^i - l^i$, adding and subtracting α, and defining $L(\tau) \equiv \alpha - (h')^{-1}(1-\tau)$ we obtain the desired expression.
- A higher tax rate on labor income distorts the labor-leisure choice and induces the consumer to work less. This will be the cost of redistributive taxation in this model.

Application: Redistributive Taxation III

- Let l denote average labor supply. Since the average of α^i is α, we have $l = L(\tau)$. The government budget constraint can therefore be written:

$$T \leq \tau l \equiv \tau L(\tau).$$

Application: Redistributive Taxation IV

- Let $U(\tau;\alpha^i)$ be utility for α^i from tax τ with T determined as residual. By straightforward substitution into the individual utility function, we can express the policy preferences of individual i as

$$U(\tau;\alpha^i) \equiv L(\tau) + h(\alpha - L(\tau)) + (1-\tau)(\alpha^i - \alpha). \tag{1}$$

- Are the preferences represented by (1) single-peaked?
- The answer depends on the shape of the average labor supply function $L(\tau)$. By putting enough structure on dysfunction, we could ensure that $U(\tau;\alpha^i)$ is strictly concave or quasi concave, thus satisfying single-peakedness. However, this function could be sufficiently convex that $U(\tau;\alpha^i)$ could have multiple peaks (multiple local maxima). As a result, preferences may not be single peaked.
- But it is straightforward to verify that (1) satisfies the single-crossing property.

Single Crossing: Recap

Definition

Consider an ordered policy space $\mathcal{P}$ and also order voters according to their α_i's. Then, the preferences of voters satisfy the **single-crossing property** over the policy space $\mathcal{P}$ when the following statement is true:

$$\begin{aligned} \text{if } p &> p' \text{ and } \alpha_{i'} > \alpha_i, \text{ or if } p < p' \text{ and } \alpha_{i'} < \alpha_i, \text{ then} \\ U(p;\alpha_i) &> U(p';\alpha_i) \text{ implies that } U(p;\alpha_{i'}) > U(p';\alpha_{i'}). \end{aligned}$$

- Notice that while single peakedness is a property of preferences only, the single-crossing property refers to a set of preferences over a given policy space $\mathcal{P}$. It is therefore a joint property of preferences and choices.

Single Crossing versus Single Peakedness

- Single-crossing property is does not imply single-peaked preferences.

$$\begin{array}{ll} 1 & a \succ b \succ c \\ 2 & a \succ c \succ b \\ 3 & c \succ b \succ a \end{array}$$

- These preferences are not single peaked. But they satisfy single crossing
- The natural ordering is $a > b > c$:

$$\begin{aligned} \alpha &= 2\text{: } c \succ b \Longrightarrow \alpha = 3\text{: } c \succ b \\ \alpha &= 2\text{: } \begin{array}{c} a \succ c \\ a \succ b \end{array} \Longrightarrow \alpha = 1\text{: } \begin{array}{c} a \succ c \\ a \succ b \end{array}. \end{aligned}$$

Single Crossing versus Single Peakedness (continued)

- The following preferences are single peaked with the natural order $a > b > c > d$:

$$\begin{array}{ll} 1 & a \succ b \succ c \succ d \\ 2 & b \succ c \succ d \succ a \\ 3 & c \succ b \succ a \succ d \end{array}$$

- For them to satisfy single crossing, we need to adopt the same order over policies (given 1's preferences) and the order $3 > 2 > 1$ over individuals.
- But then the fact that $d \succ_2 a$ should imply that $d \succ_3 a$, which is not the case. (It is easy to verify that if one chooses the order $2 > 3 > 1$ over individuals, one would obtain a similar contradiction as $c \succ_3 b$, but $b \succ_2 c$).
- This shows that single peakedness does not ensure single crossing.
- All the same, MVT works identically under single-peaked or single-crossing preferences, so verifying single crossing is enough.

Application: Redistributive Taxation V

- Therefore, we can apply MVT, and party competition gives

$$\tau^m = \arg\max_{\tau} U(\tau; \alpha^m)$$

Hence, we have

$$L'(\tau^m)\left[1 - h'(\alpha - L(\tau^m))\right] - (\alpha^m - \alpha) \leq 0 \tag{2}$$

with complementary slackness.

- If the mean is greater than the median, as we should have for a skewed distribution of income, it must be the case that $\alpha^m - \alpha < 0$ (that is median productivity must be less than mean productivity).
- This implies that $\tau^m > 0$—otherwise, (2) would be satisfied for a negative tax rate, and we would be at a corner solution with zero taxes (unless negative tax rates, i.e., subsidies, were allowed).
- Now imagine a change in the distribution of α such that the difference between the mean and the median widens. From the above first-order condition, this'll imply that the equilibrium tax rate τ^m increases.

Application: Redistributive Taxation VI

- Lessons?
- Is redistribution working well? At some point yes, greater inequality (as measured by the distance between mean and median) leads to more redistribution.
- But is this the right way to combat inequality? Why should the preferences of the median matter?
- Moreover, some also think that greater inequality in this model leads to greater "inefficiency" of policy—more distortionary taxation.
- Why is this? The reason is only weakly related to the logic of redistribution, but more to the technical assumptions that have been made.
- In order to obtain single-peaked/single-crossing preferences, we had to restrict policy to a single dimensional object, the linear tax rate.

Application: Redistributive Taxation VII

- Moreover, is this "inefficiency" the same as *Pareto suboptimality?*
- Imagine, instead, that different taxes can be applied to different people. Then, redistribution does not necessitate distortionary taxation. But in this case, preferences will clearly be non-single-peaked—agent i particularly dislikes policies that tax him a lot, and likes policies that tax agents j and k a lot, where as agent j likes policies that tax i and k a lot, etc.

Inequality and Redistribution

- One interpretation of the previous result is that greater inequality should lead to greater redistribution.
- Despite these claims in the literature, however, there is no such unambiguous prediction.
- More importantly, there is no empirical evidence that greater inequality leads to more distribution.
- In fact, why many highly unequal societies do not adopt more redistributive policies will be one of the teams we will investigate when we come to understanding the nature of institutions

Inequality and Redistribution in the MVT Models

- Consider the previous model with the mean greater than the median
- Consider the following redistribution: take money from everybody below the median and redistribute to everybody above the median in a way that leaves all voters' ranking the same.
- This is a mean preserving spread and thus increases inequality.
- But the median has become richer relative to the mean and thus there is less redistribution.
- The gap between the mean and the median generally has little to do with inequality (except for the log normal distribution).

Understanding Nonexistence

- Game theoretically, the Condorcet paradox is not about "cycling" , but nonexistence of pure strategy equilibria.
- **Example:** three (groups of) voters, $i = 1, 2, 3$ of equal size with strictly increasing preferences

$$U(p) = u\left(p^i\right),$$

where $p = \left(p^1, p^2, p^3\right)$, with $\sum_{i=1}^{3} p^i = 1$.
- A policy will be the winner if it gets votes from 2 agents.
- Now take a winning policy (p_1, p_2, p_3) where without any loss of generality suppose that $p_1 > 0$.
- Then the following policy will always beat this winning policy $(p_1 - 2\varepsilon, p_2 + \varepsilon, p_3 + \varepsilon)$, proving that there will always be cycling.
- Therefore, no pure strategy Nash equilibrium.
- Intuition: viewed as a cooperative game, this has an empty core.

Probablilistic Voting: Main Idea

- In the above example, it appears that the discontinuity of best responses in policies is important in nonexistence.
- The main idea of probabilistic voting is to "smooth" best responses in order to get existence.
- Intuitively, there are ideological and non-policy factors, so that a small advantage due to policies will not sway all voters.

Probablilistic Voting: Introduction

- G distinct groups, with a continuum of voters within each group having the same economic characteristics and preferences.
- Electoral competition between two parties, A and B, that are "non-ideological" (only care about coming to power; is this important?).
- π_P^g:fraction of voters in group g voting for party $P = A, B$, and
- λ^g: share of voters in group g. Then expected vote share of party P is

$$\pi_P = \sum_{g=1}^{G} \lambda^g \pi_P^g.$$

- Suppose that individual i in group g has the following preferences:

$$\tilde{U}_i^g(p, P) = U^g(p) + \tilde{\sigma}_i^g(P) \tag{3}$$

  when party P comes to power, where $p \in \mathcal{P} \subset \mathbb{R}^K$.
- As usual $U^g(p)$ is the indirect utility of agents in group g
- $\tilde{\sigma}_i^g(P)$ is the non-policy benefits for i from party P coming to power.

Probablilistic Voting I

- Let us normalize $\tilde{\sigma}_i^g(A) = 0$, so that

$$\tilde{U}_i^g(p, A) = U^g(p), \text{ and } \tilde{U}_i^g(p, B) = U^g(p) + \tilde{\sigma}_i^g \tag{4}$$

- In that case, the voting behavior of individual i can be represented as

$$v_i^g(p_A, p_B) = \begin{cases} 1 \text{ if } U^g(p_A) - U^g(p_B) > \tilde{\sigma}_i^g \\ \frac{1}{2} \text{ if } U^g(p_A) - U^g(p_B) = \tilde{\sigma}_i^g \\ 0 \text{ if } U^g(p_A) - U^g(p_B) < \tilde{\sigma}_i^g \end{cases}, \tag{5}$$

- Suppose that the distribution of non-policy related benefits $\tilde{\sigma}_i^g$ for individual i in group g is given by a smooth cumulative distribution function H^g defined over $(-\infty, +\infty)$, with the associated probability density function h^g.
- The draws of $\tilde{\sigma}_i^g$ across individuals are independent.
- Consequently, the vote share of party A among members of group g is

$$\pi_A^g = H^g(U^g(p_A) - U^g(p_B)).$$

Probablilistic Voting II

- Supposed to start with that parties maximize their expected vote share.
- In this case, party A sets this policy platform p_A to maximize:

$$\pi_A = \sum_{g=1}^{G} \lambda^g H^g (U^g(p_A) - U^g(p_B)). \tag{6}$$

- Party B faces a symmetric problem and maximizes π_B, which is defined similarly. Since $\pi_B = 1 - \pi_A$, party B's problem is exactly the same as minimizing π_A.
- Equilibrium policies determined as the Nash equilibrium of a (zero-sum) game where both parties make simultaneous policy announcements to maximize their vote share.
- First-order conditions for party A

$$\sum_{g=1}^{G} \lambda^g h^g (U^g(p_A) - U^g(p_B)) DU^g(p_A) = 0,$$

Probablilistic Voting Equilibrium

- Focus first on pure strategy symmetric equilibria. Clearly in this case, we will have policy convergence with $p_A = p_B = p^*$, and thus $U^g(p_A) = U^g(p_B)$.
- Consequently, symmetric equilibrium policies, announced by both parties, must be given by

$$\sum_{g=1}^{G} \lambda^g h^g(0) DU^g(p^*) = 0. \tag{7}$$

- Therefore, the probabilistic equilibrium is given as the solution to the maximization of the following weighted utilitarian social welfare function:

$$\sum_{g=1}^{G} \chi^g \lambda^g U^g(p), \tag{8}$$

where $\chi^g \equiv h^g(0)$ are the weights that different groups receive in the social welfare function.

Weighted Social Welfare Functions

Theorem

(Probabilistic Voting Theorem) *Consider a set of policy choices $\mathcal{P}$, let $p \in \mathcal{P} \subset \mathbb{R}^K$ be a policy vector and let preferences be given by (4), with the distribution function of $\tilde{\sigma}_i^g$ as H^g. Then, if a pure strategy symmetric equilibrium exists, equilibrium policy is given by p^* that maximizes (8).*

- Most important: probabilistic voting equilibria are clearly *Pareto optimal* (given policy instruments).
- Now in fact, looking back, whenever the Median Voter Theorem applies, the equilibrium is again *Pareto optimal.*
- What does this mean?

Existence of Pure Strategy Equilibria

- However, the probability voting model is not always used properly.
- It is a good model to represent certain political interactions.
- But it is not a good model to ensure pure strategy equilibria.
- In fact, pure strategy existence requires that the matrices

$$\begin{aligned} B(0, p^*) \equiv & \sum_{g=1}^{G} \lambda^g h^g(0) D^2 U^g(p^*) \\ & \pm \sum_{g=1}^{G} \lambda^g \frac{\partial h^g(0)}{\partial x} (DU^g(p^*)) \cdot (DU^g(p^*))^T \end{aligned}$$

is negative semidefinite. (Why?)

Existence of Pure Strategy Equilibria I

- Since this is difficult to check without knowing what p^*, the following "sufficient condition" might be useful:

$$B^g(x,p) \equiv h^g(x) D^2 U^g(p) + \left| \frac{\partial h^g(x)}{\partial x} \right| (DU^g(p)) \cdot (DU^g(p))^T \tag{9}$$

is negative definite for any x and p, and each g.

Theorem

(**Pure Strategy Existence**) *Suppose that (9) holds. Then in the probabilistic voting game, a pure strategy equilibrium always exists.*

Existence of Pure Strategy Equilibria II

- But (9) is a very restrictive condition. In general satisfied only if all the H^g's uniform.
- Thus we have not solved the existence problem at all.
- To understand (9), consider the first and second order conditions in the one-dimensional policy case with first-order condition

$$\sum_{g=1}^{G} h^g \left(U^g(p_A) - U^g(p_B)\right) \frac{\partial U^g(p_A)}{\partial p} = 0$$

$$\sum_{g=1}^{G} h^g \left(U^g(p_A) - U^g(p_B)\right) \frac{\partial^2 U^g(p_A)}{\partial p^2} + \sum_{g=1}^{G} \frac{\partial h^g \left(U^g(p_A) - U^g(p_B)\right)}{\partial x} \left(\frac{\partial U^g(p_A)}{\partial p}\right)^2 < 0$$

Existence of Pure Strategy Equilibria III

- Looking at each group's utility separately, this requires

$$-\frac{\partial^2 U^g(p_A)/\partial p^2}{\left(\partial U^g(p_A)/\partial p\right)^2} > \frac{\partial h^g(U^g(p_A) - U^g(p_B))/\partial x}{h^g(U^g(p_A) - U^g(p_B))}$$

for all g.

- At the same time, this point must also be a best response for party B, so by the same arguments,

$$-\frac{\partial^2 U^g(p_B)/\partial p^2}{\left(\partial U^g(p_B)/\partial p\right)^2} > -\frac{\partial h^g(U^g(p_A) - U^g(p_B))/\partial x}{h^g(U^g(p_A) - U^g(p_B))}.$$

- A sufficient condition for both of these inequalities to be satisfied is

$$\sup_x \frac{\left|\partial h^g(x)/\partial x\right|}{h^g(x)} \leq \inf_p \left|\frac{\partial^2 U^g(p)/\partial p}{\partial U^g(p)/\partial p}\right| \text{ for all } g.$$

Existence of Mixed Strategy Equilibria

- Naturally, mixed strategy equilibria are easier to guarantee (for example, they are immediate from Glicksberg's Theorem)

Theorem

*(***Mixed Strategy Existence)** *In the probabilistic voting game, a mixed strategy equilibrium always exists.*

- But do these equilibria have the same features as the canonical probabilistic voting equilibria?
- What about other pure strategy equilibria? In general, these may exist, and they would not have policy convergence. But they are difficult to characterize except in special cases (because they are inherently asymmetric).

What Happens Now with Policy-Motivated Politicians?

- Let us go back to parties/politicians with policy preferences. Suppose that second-order conditions hold, so that without policy-motivated politicians we would have a unique symmetric equilibrium.
- What happens now?

Theorem

Suppose we are in is setup with probabilistic model and the two parties have their own policy platforms, one to the left of the median the other one to the right of the median. Then there will be no policy convergence. Moreover, the stronger our the policy preferences of the right [left] candidate, the further to the left [right] will the left [right] candidate go.

- Why?

Voting for Being Pivotal

- Suppose that voters are strategic — they vote because they think they may be pivotal and are "hyper rational" so that they can understand the likelihood of being so.
- If we have a model of pure redistributive politics with two options, then each voter will vote for the option that maximizes his or her utility (with the usual arguments after ruling out weakly dominated strategies).
- But what if there is also a "common interest" element?
- In this case, each voter would like to maximize his or her utility, but this involves taking into account when he or she will be pivotal conditional on the state. Similar to common value auctions.
- Is voting likely to work well in this case?

The Condorcet Jury Theorem

- The first person to think about such issues was again Condorcet.
- Condorcet reasoned about the jury problem, where all jurors have the same interests, and would like to convict a defendant if he is guilty.
- But each has incomplete information (say a signal about the underlying state of nature).
- Condorcet reasoned that if they all pool their information — say by voting *sincerely* — then with a sufficiently large jury, the law of large numbers will kick in and the dispersed information of the jurors will be well aggregated.
- So voting acts as a good way of information aggregation.
- This point was picked up about a century later by Francis Galton, who developed the idea of the "wisdom of the crowd" and provided fascinating evidence consistent with it.

A Modern Jury Problem

- But let's dig a little bit deeper into this (following Fedderson and Pesendorfer, 1998).
- There are n jury members who have to decide whether to convict a defendant.
- There are no conflicts of interest — all jury members would like to convict the defendant if he is guilty, denoted by the underlying state $\theta = G$, but not if he is innocent, $\theta = I$.
- Each jury starts with a common prior that the defendant is guilty with probability $\pi \in (0, 1)$.
- Then receives a signal $s = \{g, i\}$ (for example, from their reading of the evidence presented at the trial). Suppose that the signals are conditionally independent and identically distributed and satisfy

$$\Pr(s = g|\theta = G) = p \text{ and } \Pr(s = i|\theta = I) = q; \; q, p > 0.5 \; .$$

Unanimity

- The key assumption is that the jury requires unanimity to reach the verdict of $x = G$.
- Let the vote of juror j be denoted by $v_j \in \{g, i\}$. Then $x = G$ if $v_j = g$ for all j.
- Suppose also that each member j of the group has the following payoff:

$$u_j(x, \theta) = \begin{cases} 0 \text{ if } x = \theta \\ -z \text{ if } x = G \text{ and } \theta = I \\ -(1 - z) \text{ if } x = I \text{ and } \theta = G \end{cases}$$

  where $z \in [0; 1]$.

- This in particular implies that convicting an innocent defendant has a higher negative payoff when z is greater (leading to more conservative decisions).

Best Responses

- When will a juror vote to convict?
- Suppose first that the juror expects not to be *pivotal* — meaning that her vote doesn't matter. This will in particular happen when other jurors have already voted to acquit (since the jury requires unanimity to convict). In such cases her vote doesn't matter, so voting $v = G$ has no payoff implications.
- Instead, her vote matters (if and only) if she is pivotal, meaning that all $n - 1$ other jurors have voted to convict.
- In this case, she would like to induce a collective decision (a jury verdict) such that

$$x = I \text{ if } \Pr(\theta = G | \text{information set}) \leq z.$$

- This simply says that given the costs of convicting an innocent, she would only like to convict the defendant if the probability that he is guilty is greater than z.

Optimal Conviction

- To simplify the discussion, let's assume that

$$\Pr(\theta = G | s_j = g \text{ for all } j) = \frac{1}{1 + \left(\frac{1-q}{p}\right)^n \frac{1-\pi}{\pi}} > z \qquad (10)$$

so that when all information is against the defendant and if jurors had access to this information, they would be confident enough to convict him.

Sincere Voting

- Let us now focus on the case where all jurors both "sincerely" and consider the problem of juror 1 who has received signal $s_1 = i$.
- The key objects we need to compute is $P_1 = \Pr(\theta = G | s_j = g \text{ for all } j \neq 1 \text{ and } s_1 = i)$. Why?
- Under sincere voting, this probability is

$$P_1 = \frac{1}{1 + \frac{q}{1-p}\left(\frac{1-q}{p}\right)^{n-1}\frac{1-\pi}{\pi}}.$$

- Does sincere voting make sense?
- First suppose that $P_1 < z$, then together with our above assumption, this condition ensures that sincere voting is an equilibrium (and in some sense the jury system works well). Why?
- But this condition is unlikely to hold together with (10). Now suppose that $P_1 > z$. What happens?

Bayesian Nash Equilibrium

- Let us now understand how the Bayesian-Nash equilibrium works when $P_1 > z$. (We note that this will always be the case when n is large. Is 12 large?).
- Then sincere voting is not an equilibrium.
- But clearly, voting to convict always cannot be in equilibrium either.
- The Bayesian-Nash equilibrium will then be in mixed strategies. In particular, suppose that $v_j = g$ if $s_j = g$, but also

$$v_j = g \text{ with probability } \gamma \text{ if } s_j = i.$$

- For such an equilibrium, we need each juror to be indifferent between voting guilty and innocent when they receive $s_j = i$. In other words,

$$\tilde{P}_1 = \Pr(\theta = G | v_j = g \text{ for all } j \neq 1 \text{ and } s_1 = i) = z.$$

Bayesian Nash Equilibrium in the Limit

- It can be shown (as it is intuitive) that as n increases, the probability of convicting the defendant converges increases, and as $n \longrightarrow \infty$, it converges to a positive number.
- Thus large juries will over-convict — they also convict the guilty with probability 1.
- Why?
- Essentially, each juror finds it optimal to rely on the implicit information that if her vote is pivotal, it must be that others have voted to convict, and that's pretty good evidence that the defendant is guilty.
- Put differently, no one wants to be a contrarian and acquit when others are voting to convict.

Lessons

- Voting in common-interest in complete information situations will be very different than what we have seen so far.
- If voters are "hyper-rational" to be able to make such inferences, they will have a tendency to distort their information (thus not engage in "sincere voting").
- But this may also involve major inefficiencies, very different from Condorcet's Jury Theorem.
- Does this mean voting is always a very bad way of aggregating information? Well, yes and no.

A Model of a Large Election

- Feddersen and Pesendorfer (1996) consider the following environment.
- There our two states of nature, $\theta = \{0, 1\}$, and two policy choices of candidates, $x \in \{0, 1\}$.
- There are three types of voters, denoted by elements of the type space $T = \{0, 1, i\}$.
- The first two are committed voters and will always choose $x = 0$ or $x = 1$ either because of distributional or ideological reasons.
- The last one designates "independent" voters, which we normally think as the "swing voters". These independents have preferences given by

$$U_i(x, \theta) = -\mathbb{I}(x \neq \theta),$$

  where $\mathbb{I}(x \neq \theta)$ is the indicator function for the position of the candidate from being different than the state of nature.
- This implies that the voters received negative utility if the "wrong" candidate is elected.

A Common Value Model (continued)

- A candidate (policy) that obtains an absolute majority is chosen. If both options obtain the same number of votes, then one of them is chosen at random.
- Let us suppose, without loss of any generality, that the prior probability that the true state is $\theta = 0$ is $\alpha \leq 1/2$, so that state $\theta = 1$ is more likely ex ante.
- To make the model work, there needs to be some uncertainty about the preferences of other voters. One way to introduce this is to suppose that how many other voters there are (meaning how many other voters could potentially turn out to vote) and what fractions of those will be committed types are stochastically generated. (This is the assumption first developed in Myerson and Weber, 1993).

Uncertainty

- Suppose, in particular, that the total number of voters is determined by Nature taking $N+1$ independent draws from a potentially large pool of voters.
- At each draw, an actual voter is selected with probability $1-p_\phi$. This implies that the number of voters is a stochastic variable with the binomial distribution with parameters $\left(N+1, 1-p_\phi\right)$.
- Conditional on being selected, an agent is independent with probability $p_i / \left(1-p_\phi\right)$, is committed to $x=0$ with probability $p_0 / \left(1-p_\phi\right)$, and is committed to $x=1$ with probability $p_1 / \left(1-p_\phi\right)$.
- Therefore, the numbers of voters of different types also follow binomial distributions.

Uncertainty (continued)

- The probability vector (p_ϕ, p_i, p_0, p_1), like preferences and the prior probability α, is common knowledge.
- Finally, each agent knows her type and also receives a signal $s \in S = \{0, 1, \phi\}$, where the first two entries designate the actual state, i.e., $\theta = 0$ or $\theta = 1$, so that conditional on receiving the signal values the agent will know the underlying state for sure.
- The last entry means that the agent receives no relevant information and this event has probability q.
- This formulation implies that some voters will potentially be fully informed, but because all events are stochastic, whether there is indeed such an agent in the population or how many of them there are relative to committed types is not known by any of the voters.
- Voting truthfully is not necessarily optimal for independents. In fact they may prefer to abstain rather than vote according to their information (priors or some other source of signals that are not certain).

Strategies

- A pure strategy here is simply

$$\sigma : T \times S \to [\phi, 0, 1],$$

where ϕ denotes abstention.

- Clearly, $\sigma(0, \cdot) = 0$ and $\sigma(1, \cdot) = 1$ (for committed voters).
- Moreover, it is also clear that $\sigma(i, z) = z$ for $z \in \{0, 1\}$, meaning that independent informed voters will vote according to their (certain) posterior.
- This implies that we can simply focus on the decisions by uninformed independent voters, denoted by

$$\tau = \left(\tau_0, \tau_1, \tau_\phi\right),$$

which correspond to the probabilities that they will vote for $x = 0$, $x = 1$ and abstain, respectively. Recall that though "uninformed," these voters have posteriors that are not equal to $1/2$, thus have relevant information.

Swing Voter's Curse

- The key observation in the analysis of this model is that, as in the jury problem, an individual should only care about his or her vote conditional on being pivotal.
- Since they do not obtain direct utility from their votes and only care about the outcome, their votes when there is a clear majority for one or the other outcome are irrelevant.
- But this implies that one has to **condition** on a situation in which one is pivotal in a large election.
- This happens (in the unlikely event) where either an equal number of agents have voted for each choice, or one of the two choices is winning with only one vote.

Swing Voter's Curse (continued)

- This intuition is sufficient to establish the following proposition, which captures the idea of the "**swing voter's curse**".
- Let $U(x, \tau)$ be the expected utility of an uninformed independent agent to choose $x \in \{0, 1, \phi\}$, when all other independents are using (symmetric) mixed strategies given by τ.

Proposition

Suppose that $p_\phi > 0$, $q > 0$ and that N is greater than 2 and even. Then if $U(1, \tau) = U(0, \tau)$, then all uninformed independent voters abstain.

Intuition

- If $U(1,\tau) = U(0,\tau)$, meaning that an uninformed voter is indifferent between voting for either candidate (policy), then he or she must prefer to abstain.
- By continuity, we could also show that if $|U(1,\tau) - U(0,\tau)| < \varepsilon$ for ε sufficiently small, then the same conclusion will apply. This is despite the fact that uninformed voters actually have relevant information, because the prior α can be arbitrarily small.
- Intuitively, when a voter expects the same utility from the two options available to him or her, then abstaining and leaving the decision to another voter who is more likely to be informed is better.
- This is despite the fact that the voter may be leaving the decision to a committed type.
- Different from the implications of models in which swing voters are "powerful".

Implications

- The implication is that useful information will be lost in the elections, and this is the essence of the "swing voter's curse".
- Nevertheless, Feddersen and Pesendorfer also show that in large elections information still aggregates in the sense that the correct choice is made with arbitrarily high probability. In particular:

Proposition

Suppose that $p_\phi > 0$, $q > 0$ and $p_i \neq |p_1 - p_0|$, then for every $\varepsilon > 0$, there exists $\bar{N}$ such that for $N > \bar{N}$, the probability that the correct candidate gets elected is greater than $1 - \varepsilon$.

- The idea of this result is that as the size of the electorate becomes large, uninformed independents mix between the "disadvantaged" candidate and abstaining, in such a way that informed independents become pivotal with very high probability.

Discussion

- Results depend on "hyper rational voters". Is this realistic?
- On the other hand, the resulting voting rule may be "simple": abstain if you do not have strong information. But this conclusion is still follows from a complicated reasoning and sometimes mixed strategies are necessary.
- How to interpret the result that the correct action will be taken in large elections?

Any Voter's Curse

- But if voters are strategic in this fashion and vote just to be pivotal, turnout will be extremely low with even trivial costs of voting.
- Turnout has to be low in particular in order to make each voter be pivotal with a sufficiently high probability.
- No way of explaining turnout rates of 20 or 30% in large elections (let alone 60 or 70%).

Evidence?

- We will discuss evidence in the next lecture.
- But it is worth mentioning the work by Battaglini, Morton and Palfrey (2008, 2010), which looks at voting behavior in reasonable-sized lab experiments with common values (as with the model here).
- They find support for two of the key features here:
    - Swing voter's curse: abstention by low information independent voters.
    - Swing voter's cunning strategy: they mix in a way to encourage more informed independence to be pivotal (and this cunning strategy is stronger when there is greater imbalance between committed voters as theory would suggest).