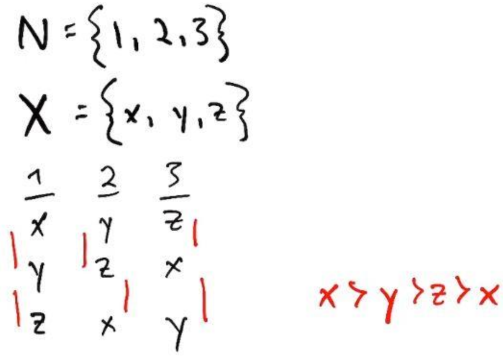
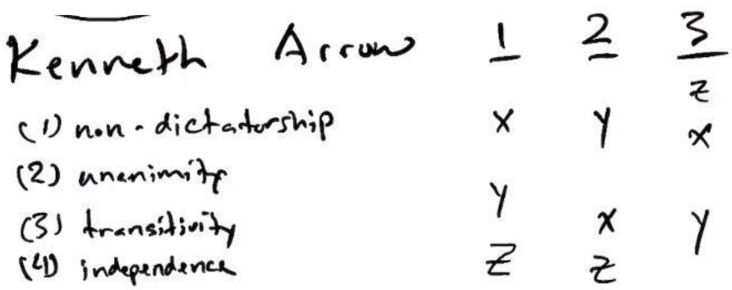
Week 1

Condorcet's Paradox: collective preferences can cycle even if individual preferences don't, as long as we're choosing among 3 or more alternatives.



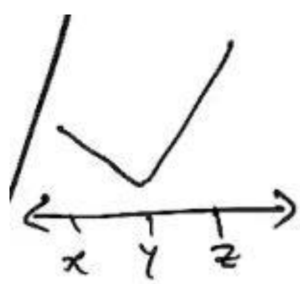
Arrow's Impossibility Theorem: it is impossible for a decision rule to satisfy the following four conditions when choosing among 3 or more alternatives: 1) non-dictatorship, 2) unanimity, 3) transitivity, 4) independence (+5: unrestricted domain). Implications are discussed in Ingham 2019 – Public Choice. Sean: it implies a conceptual problem with "the will of the people".



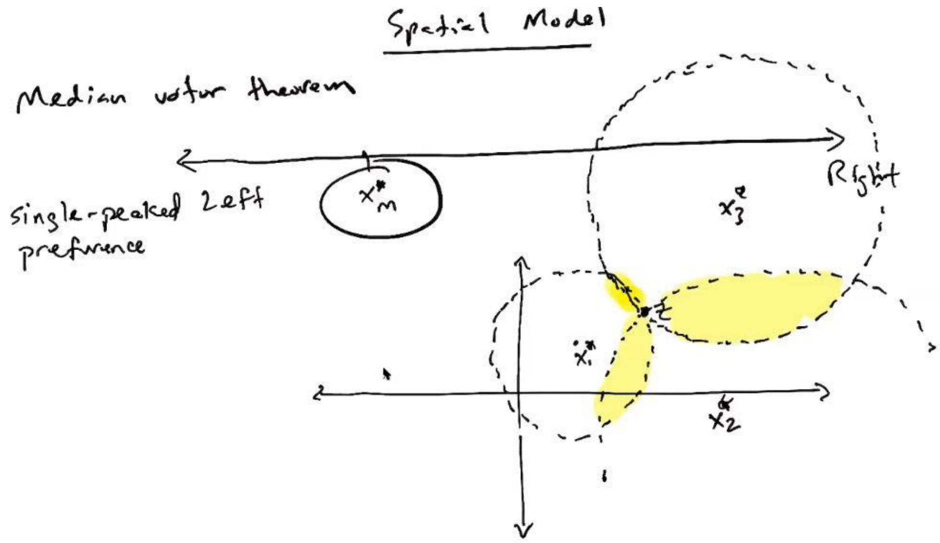
Agenda-setting is thus crucially important. *Sean: a two-party system (first-past-the-post) could be seen as an agenda-setting mechanism.*

Single-peaked preferences: part of the median voter theorem/Hotelling-Downs model. Preferences peak at x* and descend monotonically on either side. *Sean: this depends on the assumption that all of politics can be understood on a single spectrum/dimension, such as left-right, or economic-cultural.*

*Isa's question: single-peaked preferences violate the (implicit) assumption of unrestricted domain in Arrow's theorem. For example, if X and Z are both preferred to Y, we get two peaks (see below).

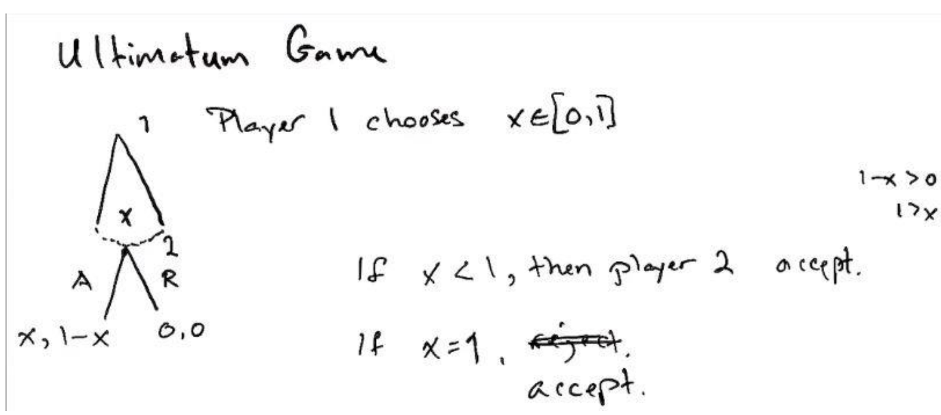


Spatial Model: If we extend this idea to a two-dimensional space, the indifference curves become circles. Note that various points (yellow) are preferred by a majority to point Z. This is a generalization of Condorcet's Paradox – for any point Z, there must be at least some shaded region that would have been preferred to it. What are we to make of majority rule under these circumstances? Sean: if we're going to institute majority rule, the reason can't be "to get a policy that the majority wants" because for any policy Z, there is necessarily a basket of other policies (in yellow) that will be preferred to it by a majority. We can't simply take the majority on one dimension, then the majority on a second dimension.



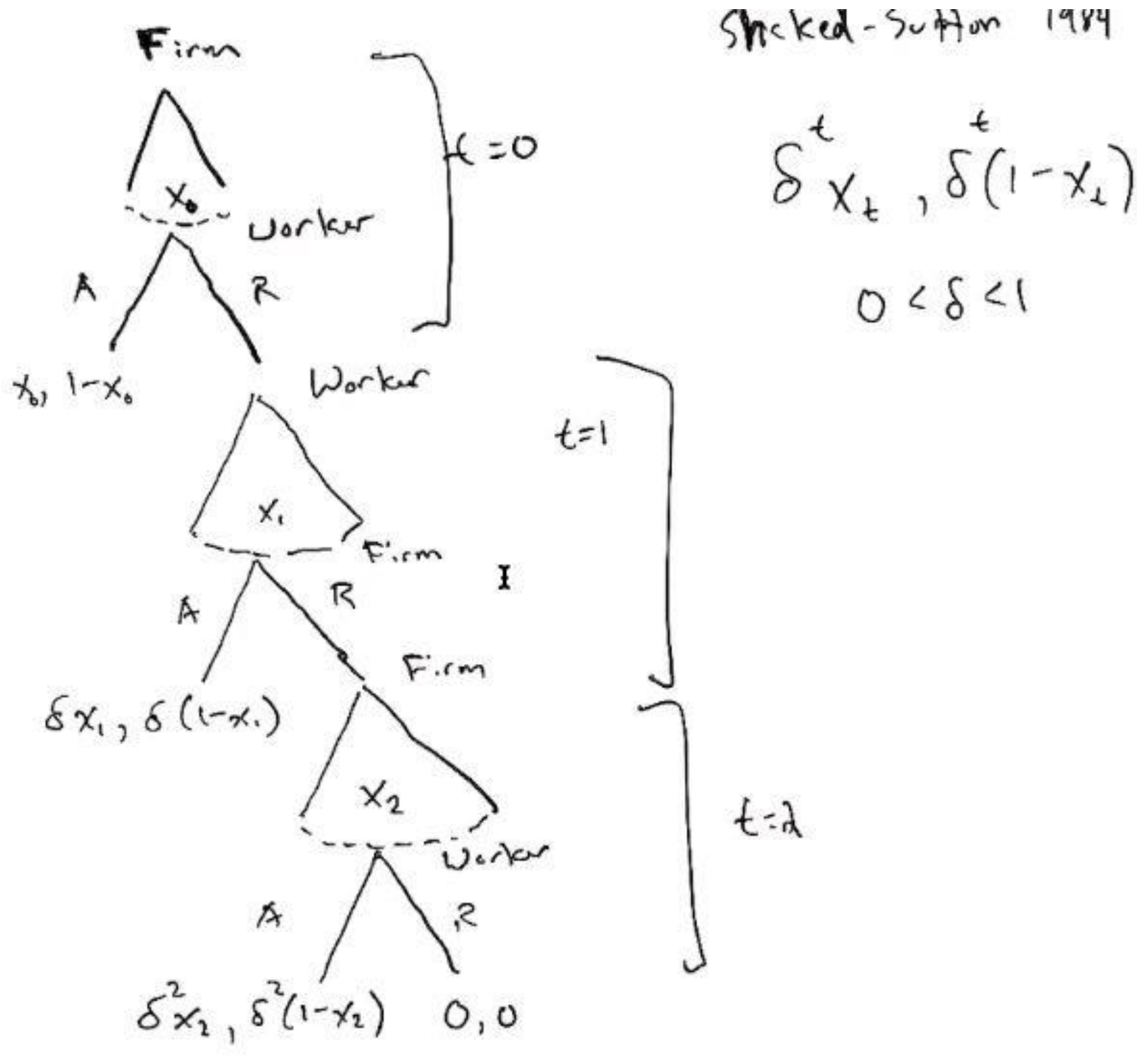
Baron and Ferejohn (reading for Week 2) are responding to this way of understanding majority rule. They ask how a legislature will allocate a surplus given these facts about majority rule. They think that we will need to say more about the institutions underpinning the process.

The Ultimatum Game (Rubinstein 1982): Since we can't use backwards induction (b/c there are an infinite number of divisions), we have to think more abstractly. Nash equilibrium can't help us because we can support any pair of strategies as mutual best responses (Tadelis p.223). Subgame perfection can get us closer to the answer. We know that if Player 1 proposes to keep any x less than 1, Player 2 will accept because something is better than nothing. If P1 proposes 0, P2 will be indifferent between accepting or rejecting. The unique SPE is therefore for P1 to offer x=1 and for P2 to accept (Tadelis p.223).

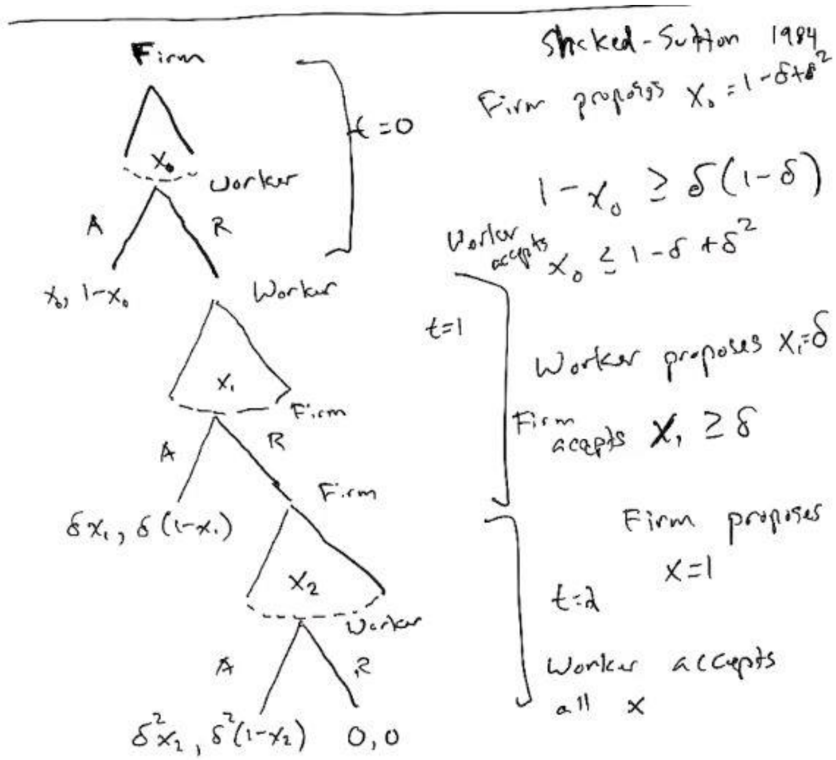


Sean: there are limits to the possibility of a rational best response. In some circumstances, our best response will be undefined. Consider a case where you can take any amount of money from a pile but must leave some behind (assuming infinite divisibility). There's no way to behave rationally in this case because there's always a smaller amount you could have chosen. Me: this recalls Zeno's paradox.

The Ultimatum Game as a bargain between management and labor (Shaked and Sutton 1984): Game extends over multiple periods, payoffs are discounted (δ). Taking these into account, if a proposal is accepted in period t, then the payoffs are ($\delta^t x_t$, $\delta(1-x_t)$).



The unique SPE of this game is for the firm (p1) to propose $x_0 = 1-\delta+\delta^2$. This is because the worker (p2) will accept any $x_0 \leq 1-\delta+\delta^2$ (we got here via backwards induction from the last stage). The proof is in the next figure.



The limit case: we know that $x_0 = (1+\delta^t)/(1+\delta)$. What happens as we increase the number of periods (as t goes to infinity)? A limit is the point at which a sequence converges. As t increases the limit of δ will be 0. As the game gets longer, the first period proposal converges to x0 → 1/(1-δ), which works out (in the limit) to ½. This makes sense, because the first player's advantage is diminishing as we add future periods.

$$\boxed{\lim_{T\to\infty} \frac{1+\delta^T}{1+\delta}} = \lim_{T\to\infty} \left(\frac{1}{1+\delta}\right) + \frac{\delta^T}{1+\delta} = \frac{1}{1+\delta} + \lim_{T\to\infty} \frac{\delta^T}{1+\delta}$$

For any $\varepsilon > 0$, there is a N such that if $n \geq N$, $|a_n - a| < \varepsilon$

$a_1, a_2, a_3, \ldots$

$$\boxed{a = \lim_{n\to\infty} a_n}$$

$$x_0 \to \frac{1}{1+\delta}$$

$a_n \to a$

$b_n \to b$

$$\boxed{a_n + b_n \to a + b}$$

$$\boxed{a_n b_n \to ab}$$

$$= \lim_{T\to\infty} \delta^T \frac{1}{1+\delta} = 0 \cdot \frac{1}{1+\delta} = 0$$

$$\delta^T \to 0 \quad 0 < \delta < 1$$

$$\lim_{\delta\to 1} \frac{1}{1+\delta} = \frac{1}{2}$$

$$\boxed{\frac{1}{2}}$$

The limit case can apply even if it's not actually reached on the path of play. If the game can <u>potentially</u> go on forever, any SPE will involve an offer of ½ that is made (and accepted) in the first period.

$$x_0 \rightarrow \frac{1}{1+\delta}$$

$$\boxed{\frac{1}{2}}$$

$$\lim_{\delta \to 1} \frac{1}{1+\delta} = \frac{1}{2}$$

Game with infinitely-many periods. We ask, suppose we have an SPE, consider the subgame beginning in the final round. Ask what the supremum would be of all these equilibrium payoffs. Supremum (M): the least upper bound of a set. This value (M) will also turn out to be the continuation value to the firm (if they decide to continue the game, the best possible payoff they can receive is defined as M).

$$M = \sup \left\{ v_f \in \mathbb{R} \;\middle|\; v_f \text{ is firm's payoff in a SPE to the game beginning in period 2} \right\}$$

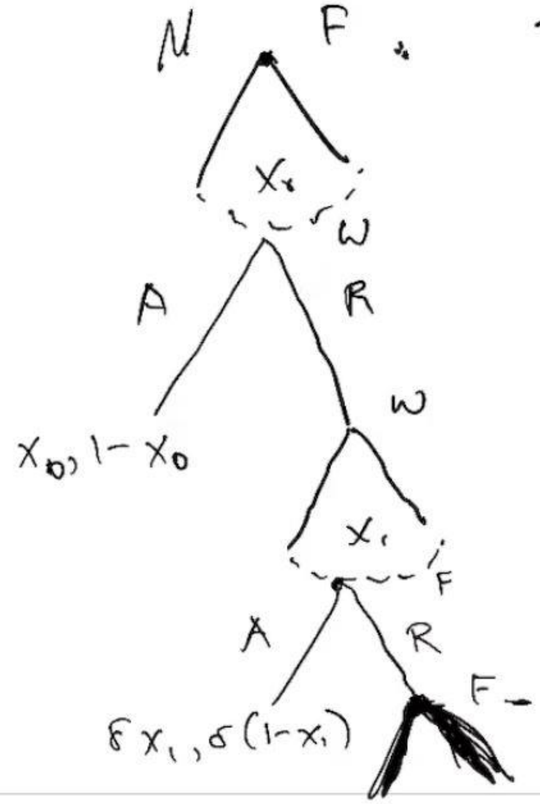
M is the continuation value to the firm

Since subgames of this infinite game are themselves infinite, the value of starting the game must be the same at any point (for subgames or for the game as a whole). That value is:

$$M = 1 - \delta + \delta^2 M$$

$$M(1-\delta^2) = 1 - \delta$$

$$M = \frac{1-\delta}{1-\delta^2} \boxed{= \frac{1}{1+\delta}}$$



$\boxed{x_0 = 1-\delta+\delta^2 M}$

period 0

$1-x_0 \geq \delta(1-\delta M)$

$\boxed{x_0 \leq 1-\delta+\delta^2 M}$

period 1

W proposes $x_1 = \delta M$

F accepts $x_1 \geq \delta M$

period 2 M

Week 2

The "core" of the majority preference relation. Majority preference relation is a binary relation defined as:

$X = \{x, y, \ldots\}$, $|X| \geq 3$

$N = \{1, 2, \ldots, n\}$, n odd

For each $i \in N$, $\succsim_i$

Majority preference relation is a binary relation $\succsim$

$x \succsim y$ iff $\left|\{i \in N : x \succsim_i y\}\right| \geq \frac{n+1}{2}$

The core consists of all elements that are as good as any other elements according to the majority preference relation. And the core is defined as:

$$\text{core}(\succsim) := \{x \in X : x \succsim y \text{ for all } y \in X\}$$

"Condorcet winner"

"Condorcet Winner" is another way of defining the elements in the core. Each of these is preferred to all others.

Problem 1a from the Week 1 problem set:

If not $x \succsim y$, then $\boxed{y \succ x.}$

Crucial to note that a binary preference relation means that:

$X = \{(x_1, \ldots, x_n) \in [0,1]^n : x_1 + x_2 + \ldots + x_n = 1\}$

For each $i \in N$, $x \succsim_i y$ iff $x_i \geq y_i$.

$\text{core}(\succsim) = \emptyset$

Choose any $x \in X$. Without loss of generality, assume $x_1 > 0$. Now define y as $\left(0, x_2 + \frac{x_1}{n-1}, \ldots, x_n + \frac{x_1}{n-1}\right)$. $y \in X$. And $y \succ_i x$ for all $i \neq 1$. Thus $y \succ x$. Thus, there is no $x \in X$ such that $x \succsim y$ for all $y \in X$. Thus, $\text{core}(\succsim) = \emptyset$. $\square$

Problem 1b

For any alternatives x and y, we can construct a path leading from the first to the second, where every step in the path is preferred by a majority to the previous step.

$y = (1/3, 1/3, 1/3)$ $\qquad 0 < \delta < 1/3$

$x = (0, \delta, 1-\delta)$

Prove: there are alternatives $w, z \in X$ s.t. $x \succ z \succ w \succ y$.

Fix $0 < \varepsilon < \delta$. Let $w := (2/3 - \varepsilon, 1/3 + \varepsilon, 0)$ and $z := (1-\varepsilon, 0, \varepsilon)$.

Then $x \succ z \succ w \succ y$.

This problem demonstrates how important control over the agenda and voting procedures are. It's always possible to construct a sequence that arbitrarily benefits particular players. This is a special case of a more general idea, which is that in any spatial model where the set of alternatives is some subset of a multidimensional space, not only will the core of the majority preference relation be empty, but it will usually be possible to construct a path from any distribution to another distribution such that each step in the path is preferred by a majority. See McKelvey's "chaos" theorems.

Shepsle (and others) imagine institutions as the sequencing of choice options – who gets to decide when. Sean calls this "the naïve view". This view takes democratic institutions to be those permitting "the majority" to rule, meaning that it can choose laws and policies reflecting its will. The key problem is that the naïve view ignores institutions. On this view, institutional details don't affect our generalizations about the kinds of policies that democracies deliver, though they might affect our views on whether institutions or regimes *qualify* as democracies. The naïve view is concerned with outcomes, the institutionalist view is concerned with procedure. Sean: the results of social choice theory make the naïve view indefensible.

"Naive" view

Democracy = institutions that empower majority to choose policies that reflect the will of the majority.

"Institutionalist" view

Democracy = for any policy, any majority, there are actions available to the members of that majority that result in that policy.

An alternative view: we can't define democracy as policies reflecting the will of the majority, because that phrase doesn't refer to anything. So democracy must be something else. The institutionalist view sees it as merely procedural – the rules of the game allowing actions at different times leading to different decisions. For any policy, there must be actions available to the members of a majority that result in the implementation of that policy.

Example with amendment rules (open/closed) – a proponent of the naïve view might view this distinction as a determinant of how democratic the state is, but an institutionalist would interpret this as two possible types of democracy.

Notice that the Hotelling-Downs model and the median voter theorem are perilously close to the naïve view. Acemoglu and Robinson define democracy as the opportunity for a median voter to implement policies – this is also similar to the naïve view. "Assuming that there will always be a median voter is only slightly less objectionable than assuming that we can always refer to "the will of the majority". In general, if the core of the majority preference relation is empty then there is no median voter, because if there is a median voter then that voter's ideal point should be the only element of the core.

The Baron and Ferejohn Model – Finite

Note that Baron and Ferejohn restrict their analysis to weakly undominated strategies. Note also that at an outcome where no player is casting a pivotal vote, no single player can profitably deviate because the outcome doesn't depend on their vote, so each way of voting is a best response *no matter their preferences*. So only pivotal votes have to be cast according to preferences. Voting contrary to preference is weakly dominated in the voting game, because individuals can never do better by voting contrary to preference. This is why they're only considering weakly undominated strategies – only the cases where individuals have an incentive to cast votes reflecting actual preferences.

In any SPE w/ $\boxed{\text{weakly undominated strategies,}}$ period-2 proposer proposes keeping everything, and at least majority vote in favor.

$$\frac{1}{n} \cdot 1 + \frac{n-1}{n} \cdot 0 = 1/n$$

votes for x if $x_i \underset{\geq}{>} \delta/n$

first-period proposal will give $1 - \frac{n-1}{2}\frac{\delta}{n}$ to the proposer, δ/n to $\frac{n-1}{2}$ other players, and 0 to the rest

On the equilibrium path of play, in any subgame-perfect equilibrium with weakly undominated strategies, the period-2 proposer will propose keeping everything, and at least a majority will vote in favor. So in the first period, the continuation value of starting the game is 1/n (probability of being recognized) * 1 (the payoff) + n-1/n (probability of not being recognized) * 0 (the payoff) = 1/n (the continuation value). All players will vote for a first-period proposal x if x_i is greater than δ/n.

The Baron and Ferejohn Model – Infinite

Sufficiency claim – because all subgames beginning with the choice of a proposer are identical, we can use the notation v_i to indicate the common value of all players' continuation values.

$$v_i = \frac{1}{n}\left(1 - \frac{(n-1)}{2}\frac{\delta}{n}\right) + \frac{n-1}{n}\frac{1}{2}\cdot\frac{\delta}{n}$$

$$= \frac{1}{n} - \frac{n-1}{n}\frac{1}{2}\frac{\delta}{n} + \frac{n-1}{n}\frac{1}{2}\cdot\frac{\delta}{n}$$

$$= 1/n$$

$$x_i \geq \delta/n$$

$$\frac{n-1}{2}$$

$$\boxed{1 - \frac{(n-1)\delta}{2n} \geq \frac{\delta}{n}}$$

Sean's example (in the lecture notes) prove that it's possible to have equilibria where players have the same continuation values even if players aren't constructing their coalition at random.

The Baron and Ferejohn Model – Open Rule

n=3

(i) each player proposes division in which proposer gets x and the other two get $\frac{1-x}{2}$ each.

(ii) each player (if recognized as amender) moves the motion on the floor to a vote if it gives that player at least $\frac{1-x}{2}$, otherwise proposes an amendment according to (i).

(iii) vote for whichever of two proposals give them the most, vote for the second one if indifferent, and votes for a proposal that has been seconded / moved to a vote if it offers at least $\frac{1-x}{2}$.

$\left(x, \frac{1-x}{2}, \frac{1-x}{2}\right)$ $\quad \left(\frac{1-x}{2}, x, \frac{1-x}{2}\right)$ $\quad \boxed{x = \frac{1}{1+2\delta}}$ $\quad \frac{1-x}{2} = \delta x$

Proof that these strategies constitute a subgame-perfect equilibrium:

$$x = \frac{1}{1+2\delta} \qquad \boxed{\frac{1-x}{2} = \boxed{\frac{\delta}{1+2\delta}}}$$

$$\boxed{\left(1 - \frac{\delta}{1+2\delta}, \frac{\delta}{1+2\delta}, 0\right)} \qquad \left(\frac{\delta}{1+2\delta}, \frac{\delta}{1+2\delta}, \frac{1}{1+2\delta}\right)$$

$$\frac{1}{1+2\delta} \geq \frac{1}{2}\left(1 - \frac{\delta}{1+2\delta}\right) + \frac{1}{2}\left(\frac{\delta^2}{1+2\delta}\right)$$

$$1 \geq \delta(1+\delta)$$

Week 3

Hobbes makes an argument for radical democracy. Sean sets up his interest in this material by referring to the debate between Hobbes and Locke as one between radical democrats and constitutionalists. Thesis of Absolute Sovereignty: there must (at some level) be a sovereign with absolute power, either an individual or a committee. Sean emphasizes that Hobbes' social contract is between individuals, and that because the sovereign is not a party to the contract it cannot violate it.

Consequently they that have already Instituted a Common-wealth, being thereby bound by Covenant, to own the Actions, and Judgements of one, cannot lawfully make a new Covenant, amongst themselves, to be obedient to any other, in any thing whatsoever, without his permission. And therefore, they that are subjects to a Monarch, cannot without his leave cast off Monarchy, and return to the confusion of a disunited Multitude; nor transferre their Person from him that beareth it, to another Man, or other Assembly of men: for they
[89] are bound, every man to every man, to Own, and be reputed Author of all, that he that already is their Soveraigne, shall do, and judge fit to be done: so that any one man dissenting, all the rest should break their Covenant made to that man, which is injustice: and they have also every man given the Soveraignty to him that beareth their Person; and therefore if they depose him, they take from him that which is his own, and so again it is injustice. Besides, if he that attempteth to depose his Soveraign, be killed, or punished by him for such attempt, he is author of his own punishment, as being by the Institution, Author of all his Sovereign shall do: And because it is injustice for a man to do any thing, for which he may be punished by his own authority, he is also upon that title, unjust. And whereas some men have pretended for their disobedience to their Soveraign, a new Covenant, made, not with men, but with God; this also is unjust: for there is no Covenant with God, but by mediation of some body that representeth Gods Person; which none doth but Gods Lieutenant, who hath the Soveraignty under God. But this pretence of Covenant with God, is so evident a lye, even in the pretenders own consciences, that it is not onely an act of an unjust, but also of a vile, and unmanly disposition.

2. Soveraigne Power cannot be forfeited.

Secondly, Because the Right of bearing the Person of them all, is given to him they make Soveraigne, by Covenant onely of one to another, and not of him to any of them; there can happen no breach of Covenant on the part of the Soveraigne; and consequently none of his Subjects, by any pretence of forfeiture, can be freed from his Subjection. That he which is made Soveraigne maketh no Covenant with his Subjects before-hand, is manifest; because either he must make it with the whole multitude, as one party to the Covenant; or he must make a severall Covenant with every man. With the whole, as one

Covenants after he hath the Soveraignty are voyd, because what act soever can be pretended by any one of them for breach thereof, is the act both of himselfe, and of all the rest, because done in the Person, and by the Right of every one of them in particular. Besides, if any one, or more of them, pretend a breach of the Covenant made by the Soveraigne at his Institution; and others, or one other of his Subjects, or himselfe alone, pretend there was no such breach, there is in this case, no Judge to decide the controversie: it returns therefore to the Sword again; and every man recovereth the right of Protecting himselfe by his own strength, contrary to the designe they had in the Institution. It is therefore in vain to grant Soveraignty by way of precedent Covenant. The opinion that any Monarch receiveth his Power by Covenant, that is to say on Condition, proceedeth from want of understanding this easie truth, that Covenants being but words, and breath, have no force to oblige, contain, constrain, or protect any man, but what it has from the publique Sword; that is, from the untyed hands of that Man, or Assembly of men that hath the Soveraignty, and whose actions are avouched by them all, and performed by the strength of them all, in him united. But when an [90]
Assembly of men is made Soveraigne; then no man imagineth any such Covenant to have past in the Institution; for no man is so dull as to say, for example, the People of *Rome*, made a Covenant with the Romans, to hold the Soveraignty on such or such conditions; which not performed, the Romans might lawfully depose the Roman People. That men see not the reason to be alike in a Monarchy, and in a Popular Government, proceedeth from the ambition of some, that are kinder to the government of an Assembly, whereof they may hope to participate, than of Monarchy, which they despair to enjoy.

3. No man can without injustice protest against the Institution of the Soveraigne declared by the major part.

Thirdly, because the major part hath by consenting voices declared a Soveraigne; he that dissented must now consent with the rest; that is, be contented to avow all the actions he shall do, or else justly be destroyed by the rest. For if he voluntarily entered into the Congregation of them that were assembled, he sufficiently declared thereby his will (and therefore tacitely covenanted) to stand to what the major part should ordayne: and therefore if he refuse to stand thereto, or make Protestation against any of their Decrees, he does contrary to his

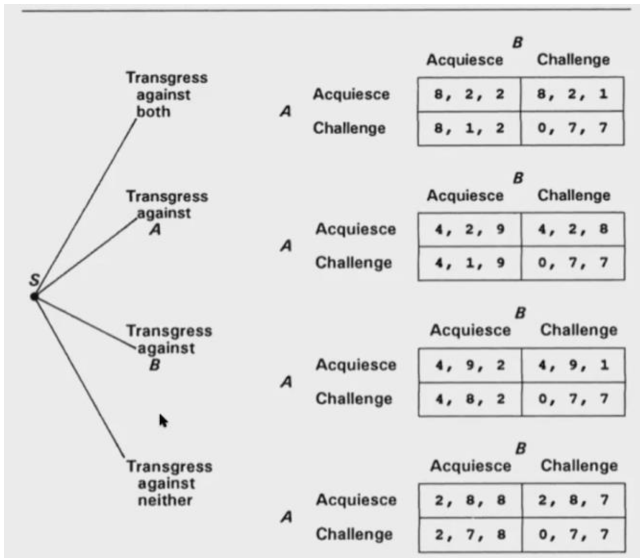
He makes much of the passage that indicates that there will be no judge to decide a controversy between one set of subjects and another (with regard to the sovereign's actions). The controversy will thus return to the state of nature. He emphasizes that radical democrats in Hobbes' tradition would not see constitutional binding or judicial review as a solution to this problem. "When political actors are constrained by political rules, it is because if they were to violate those rules, there would be some coordinated punishment activity by others, and because they are aware of this, they don't violate them."

I raised Hardin's bootstrapping problem as an analogous "impossible" case. Sean pointed out that Hobbes contradicts himself by saying that contract formation is impossible in the state of nature but then attributing the creation of the sovereign to a contractual process among citizens in the state of nature. Kevin reminded us that Lawrence emphasizes that equilibria are sticky, and that we might fall into them at random but then have good reasons for staying in them.

Sean: why should an individual expose themselves to risk by helping to enforce the sovereign's commands? Maybe because we expect others to practice 3rd-party enforcement against us. Kevin brought up Wiens' argument that the state of nature is not a prisoners' dilemma but rather a stag hunt. Sean points out another problem in Hobbes' reasoning: since we don't have to obey commands that would cause us physical harm (ch. 21), why are we obliged to practice costly (and potentially harmful) 3rd party enforcement? Since we know that they aren't actually obligated to enforce the sovereign's commands, we should expect that they won't, which should make the whole schema unravel.

Weingast's Model

Most of Sean's commentary is along the lines of problem set #2. He thinks that there's nothing in Weingast's model that justifies us in labelling certain actions taken by the sovereign in the real world as 'transgressions'. Because the payoff structure is common knowledge, that implies consensus on the very thing that Weingast was arguing varied by individual in his model – the content of the idea of a "transgression against rights".



The model either assumes that we all agree on what counts as a transgression against rights, or (if the payoffs encode satisfaction from different states of the world) makes very strong assumptions about the extent to which we understand (have perfect information about) our peers' preferences and how they would react to the sovereign's actions. Payoffs are common knowledge, and they have to represent either a consensus on what transgressions are or consensus on how transgressions would be mediated through preferences.

Sean: Weingast is trying to explain things in the real world based on how he set up his model. Kevin: The simple model characterizes the situation where the society agrees about what transgressions against rights are, and the more complex model represents a society where they disagree. Sean: that's right, but notice that it's an unsatisfying explanation for democracy to say that it happens in societies where we agree about transgressions. What causes democracy happens outside the model. Me: it's also not how Weingast substantively interprets his own model – he thinks it's about how citizens solve coordination problems.

Kevin: In the second half of the paper, Weingast draws substantive conclusions from the equilibria in his model. But because the folk theorem says that we can sustain almost any allocation in equilibrium with relatively patient discount rates, can't we just use the model to draw *any* substantive conclusions we feel like drawing? Anything can be sustained as a focal point. Sean: the model is the only part of the paper that allows us to prove Weingast wrong. This shows the use of models – they make it possible to analytically engage with the work.

Week 4 – Fearon (2011)

Fearon's model involves a continuum of citizens (i) between 0 and 1. In each period, the ruler chooses some number g_t that is between 0 and v (the total amount of surplus). Individuals observe the ruler's choice, but filtered through some random shock (epsilon, between 0 and 1). The probability that the epsilon will be 1 (i.e. that the ruler's choice will accurately be conveyed to the citizen) is alpha.

After observing these outcomes, each citizen chooses whether or not to rebel. If some fraction M_t rebels, then the probability of successful revolution is $G(M_t)$, where G is a continuous c.d.f. In the simple versions of the model, if all citizens rebel the rebellion succeeds (with probability 1).

Citizen i will incur a cost c for participating in an unsuccessful revolution, and will realize a "warm glow" benefit b for participating in a successful revolution. All citizens pay a disruption cost D, which is greater than b. Note that Fearon is assuming away the free-rider problem in collective action – citizens gain a greater benefit from participating in successful collective action than they would if it succeeded without their involvement.

- $i \in [0,1]$
- ruler $g_t \in [0,v]$
- i observes $x_{it} = g_t \varepsilon_{it}$, where $\varepsilon_{it} \in \{0,1\}$, $prob(\varepsilon_{it}=1)=\alpha$

  then i chooses rebel or not
- if a fraction M_t, then prob. of successful revolution is $G(M_t)$, where G is a continuous c.d.f.,

  $G(0)=0$, $G(1)=1$
- $b \geq 0$, $c > 0$, $\boxed{D > b}$

Citizen's payoff: $u_{it}(p_{it}, g_t, M_t) = \underbrace{g_t \varepsilon_{it}}_{x_{it}} + p_{it}\left[\underline{G(M_t)}b - (1 - G(M_t))\right] - 1(M_t)D$

Ruler's payoff: $v - g_t$

$1(M_t) = 0$ if $M_t = 0$

$1(M_t) = 1$ otherwise

First model: citizens know g_t. (Markov strategy conditioned on g_t). On the equilibrium path, leaders are never removed. The ruler can't do any better by deviating, and the citizen will definitely do worse.

Proof of Proposition 1:

$g^* = \underline{\delta v}$

i rebels iff $g_t < \delta v$

$$\sum_{t=1}^{\infty} \delta^{t-1}(v - \delta v) = \frac{v - \delta v}{1 - \delta} = v$$

$$\boxed{\underline{\alpha \delta v} > \alpha \delta v \boxed{-c}} \qquad g_t > \delta v$$

$$\alpha g_t + b - D > \alpha g_t - D \qquad \boxed{g_t < \delta v}$$

$$b > 0 \qquad M_t > 0$$

In-class example: changing the payoff function so that citizens prefer *less* government spending still preserves the equilibrium. The same strategies constitute an equilibrium even if we put a -1 in front of what the citizen observes. The ruler's choice of the governance outcome g has strategic implications for equilibrium behavior, but just as a coordinating device, not because it plays a role in the citizens' payoff functions. Fearon has set up the model to give citizens a reason to desire successful revolutions no matter the underlying circumstances. The citizen doesn't care what g is. This means there are equilibria where the ruler hams the citizens. Sean calls this a limit to moral hazard models, because if all agents are the same type (i.e. if the next ruler will face the same incentives) it's difficult to explain accountability.

$g^* = \underline{\delta v}$

i rebels iff $g_t < \delta v$

$$\sum_{t=1}^{\infty} \delta^{t-1}(v - \delta v) = \frac{v - \delta v}{1 - \delta} = v$$

$$\boxed{\underline{\alpha \delta v} > \alpha \delta v \boxed{-c}} \qquad g_t > \delta v$$

$$\alpha g_t + b - D > \alpha g_t - D \qquad \boxed{g_t < \delta v}$$

$$b > 0 \qquad M_t > 0$$

Question for Sean: You've said several times in class that it's desirable to formalize thinking into models to ensure that we aren't making any mistakes or unwarranted assumptions. I'm very sympathetic to this view. However, it seems (at least superficially) discouraging that when trying to commit their arguments to game-theoretic notation both Fearon and Weingast made what appear to be substantial mistakes that were not noticed or addressed prior to publication. I've identified several possibilities for why it might nevertheless be worthwhile to model our theoretical speculations, and I'm interested to know whether you think any of them are correct. First, we might say that the attempt itself is worthwhile. That is, in the process of coming up with their (imperfect) models, both Fearon and Weingast were able to unearth counterintuitive conclusions that enriched our understanding of the phenomena being studied. This doesn't seem empirically correct in the case of these two papers, but I can certainly imagine it occurring in other contexts. Second, we might say that, though flawed, the modeling approach is superior to the (also flawed) verbal alternative. I'm thinking of the approach taken by Schelling in Arms and Influence as an example of how else this type of speculation might be undertaken. Finally, it may be the case that the mistakes would have been present in any case, and only the authors' explicit choice of a modeling approach allows us to locate mistakes in Fearon and Weingast's thinking. I want to believe this, but because these mistakes survived peer review it's a difficult belief to sustain. This is all to say that while I'm very sympathetic to the view that models are a useful device for disciplining our thinking, that doesn't appear to be what is happening in these particular models. Far from being disciplined by his modeling assumptions, Fearon's thinking was actually led astray by them. The modeling assumption that choices by the ruler have no impact on the probability of protest appears to vitiate any conclusions that we might draw about the world as it is.

My motivation for asking this question is that when using game theory in my own work, I've found that in addition to disciplining my reasoning, it introduces the potential for new kinds of error. Specifically, the use of modeling techniques allows us to reach conclusions on the basis of our notation alone. It is difficult for me to locate a standard of proof that will allow me to know when conclusions reached on the basis of notation are in fact interpretable as conclusions about the real situation I'm trying to model. Far more often, the ostensible conclusions are in fact based on a mistaken modeling assumption. Although it's comforting to see luminaries like Fearon make similar mistakes, the frequency of these errors seems to throw the utility of the technique into question. While these techniques allow us to draw counterintuitive conclusions, in the absence of a basis for establishing correspondence between our modeling assumptions and reality it seems impossible to exit the model and draw conclusions about the real world.

Second model: citizens can’t observe g_t, but only observe their own individual outcomes.

Proof of Proposition 2: