Dynamic Games – Review

- Multi-stage games with observed actions
  - "with observed actions" requires that at the end of each stage, the players observe what everyone else has done.
  - At each stage, all players must move simultaneously
  - Examples of multi-stage games
    - Repeated prisoner's dilemma
    - Rubinstein-Stahl Bargaining
      - Time 0: P1 makes offer, P2 accepts/rejects
      - Time 1: if P2 rejects, then P2 makes offer, and P1 accepts/rejects
      - Time 2: if P1 rejects, then …
      - In this example, "stage" and "time" are not necessarily the same. Can consider a time period as consisting of two stages, and within each stage, one of the players has a trivial action.
- **One-stage deviation principle (finite horizon).** A strategy profile s is a SPE if and only if no player i can gain by deviating from s at a single history, and conforming to s thereafter, given any history.
  - Example (infinitely repeated prisoner's dilemma)

	C	D
C	11, 11	0, 12
D	12, 0	10, 10

Question: Is "both players play tit-for-tat" a SPE? (assume $\delta \in \left(\frac{1}{11}, \frac{1}{2}\right)$)

- Suppose the opponent plays tit-for-tat.
- If cooperate, get a payoff of 11
- If deviate for only one period, payoff is

$$12(1-\delta) + 0 + 12\delta^2(1-\delta) + \cdots = (1-\delta)\frac{12}{1-\delta^2} = \frac{12}{1+\delta} < 11$$

given $\delta > \frac{1}{11}$.

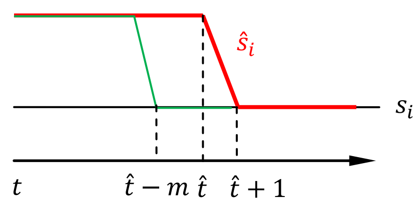
- If play D forever, $12(1-\delta) + 10\delta > 11$.
- This creates an "anomaly". But this due to a mistake: we've only checked the optimality of deviation on the equilibrium path, but we haven't checked the optimality off equilibrium paths.

  - *Proof.* (⇒) if s is SPE, then it follows trivially that no player can deviate profitably at a single history and conforming back to s thereafter.
(⇐) Suppose s satisfies the condition, but is not subgame perfect.

- Then, there exists h^t such that some $\hat{s}_i$ is a better response than s_i against s_{-i} in the subgame starting at h^t.
- Let $\hat{t}$ be the largest t' such that for some $h^{t'}$, $\hat{s}_i(h^{t'}) \neq s_i(h^{t'})$, so $\hat{s}$ and s_i match after $\hat{t}$.
- We know $\hat{t} > t$. By assumption, $\hat{t}$ is finite.
- Consider

$$\tilde{s}_i^0 = \begin{cases} \hat{s}_i & \text{before } \hat{t} \\ s_i & \text{at and after } \hat{t} \end{cases}$$

- We know that $\hat{s}_i$ is no better than $\tilde{s}_i^0$ at any history $h^{\hat{t}}$, differ only at $\hat{t}$, and $\tilde{s}_i^0 = s_i$ at such histories.



- If $\hat{t} = t + 1$, we're done, since $s_i = \tilde{s}_i^0$, starting at $\hat{t}$
- If $t < \hat{t} - 1$, then define

$$\tilde{s}_i^m = \begin{cases} \hat{s}_i & \text{before } \hat{t} - m \\ s_i & \text{at and after } \hat{t} - m \end{cases}$$

- Proceed inductively in an analogous manner.

❖ *Definition.* A game is ***continuous at infinity*** if

$$\sup_{h,\tilde{h}:h^t=\tilde{h}^t} |u_i(h) - u_i(\tilde{h})| \to 0, \qquad as\ \ t \to \infty$$

- E.g. In repeated games where payoffs are bounded and $\delta < 1$ are continuous at ∞
- Thus, for games that are continuous at infinity, the one-stage deviation principle holds for games of infinite horizons.

❖ Repeated games

- **Nash reversion folk theorem.** Suppose there exists a static NE with payoffs v^*. Then, for every $v > v^*$, there exists a $\underline{\delta} < 1$ such that $\forall \delta \in (\underline{\delta}, 1)$, $\exists$ an SPE $G(\delta)$ with payoffs $v \in V$, where V is the feasible set.
  - Sometimes, this folk theorem is useless. In the following example, where the NE payoff is $(0,1)$, there exists no feasible payoff vector to the northeast of $(0,1)$.

	L	R
U	-2, 2	1, -2
M	1, -2	-2, 2
D	0, 1	0, 1

- *Definition.* A player i's ***minmax*** payoff in a static game is

$$\underline{v}_i = \min_{\alpha_{-i}} \left[\max_{\alpha_i} u_i(\alpha_i, \alpha_{-i}) \right]$$

Folk Theorem

- **Folk Theorem (Fudenberg & Maskin 1986).** If the dimension of the feasible set is equal to the number of players, $\dim(V) = \#(i)$, then for all $v \in V$ such that $v_i > \underline{v}_i$, there exists $\underline{\delta} < 1$ such that there exists SPE of game with discount factor δ with payoffs v, for all $\delta \in (\underline{\delta}, 1)$.
  - *Proof.* For simplicity, assume that there exists a pure strategy profile a with payoff vector v. Let m_{-i}^{i} be the stage game profile played by players other than i to minmax player i.
    - *Case* 1. Suppose m_{-i}^{i} is pure for all i. Choose $v' \in \text{int}(V)$ and $\epsilon > 0$ such that
      (1) $\forall i : \underline{v}_i < v_i' < v_i$
      (2) $(v_1' + \epsilon, \dots, v_{i-1}' + \epsilon, v_i', v_{i+1}' + \epsilon, \dots, v_n' + \epsilon) \equiv v'(i) \in V$
      Assume that there exists a pure $a(i)$ implementing $v'(i)$. Choose N such that
      $$\max_a g_i(a) + N\, \underline{v}_i < \min_a g_i(a) + N v_i', \qquad \forall i$$
      where $g_i(a)$ is player i's payoff from outcome a in the stage game. In words, this condition means that, if patient enough, everyone would rather get $\min_a g_i(a)$, followed by N periods of v_i' than get $\max_a g_i(a)$, followed by N periods of $\underline{v}_i$.
      - *Phase* I. Play a. Remain in phase I unless a single *j* deviates, in which case go to phase II_j.
      - *Phase* II_j. Minmax j for *N* periods, and then go to phase III_j. If a single i deviates, go to phase II_i.
      - *Phase* III_j. Play $a(j)$, and remain in phase III_j unless a single *i* deviates, in which case go to II_i.
    - *Case* 2. Suppose m_{-i}^{i} is not pure. The idea is to make the payoffs in phase III vary slightly (i.e. by a lot less ϵ) depending on actual outcomes in phase II in such a way that the players needing to mix in m^i are indifferent between their actions. ■
- Other extensions.
  - Public and/or private observations are not perfect.

Markov Perfect Equilibrium

- ***Stochastic games***
  - States: $k \in K$
  - Action spaces: $A_i(k)$
  - Transition function: $q(k^{t+1}|k^t, a^t)$
  - Payoff: $u_i = \sum_{t=0}^{\infty} \delta^t g_i(k^t, a^t)$, where $g_i(\cdot)$ is a one-period payoff
    - The key here is that the action spaces and the payoffs depends only on the state, not depend on the entire history.
    - Assume that actually depend on k^t, i.e.
$$\forall k^t \neq k^{t'}, \exists a^t : g_i(k^t, a^t) \neq g_i\left(k^{t'}, a^t\right)$$
or
$$q(\cdot\,|k^t, a^t) \neq q\left(\cdot\,\middle|k^{t'}, a^{t'}\right)$$
  - Known history: $h^t = (k^0, a^0, k^1, a^1, \dots, k^{t-1}, a^{t-1}, k^t)$
- A ***Markov Strategy*** is a strategy σ_i such that $\sigma_i(h^t) = \sigma_i\left(\hat{h}^t\right)$ whenever $k^t = \hat{k}^t$, where
$$h^t = (k^0, a^0, k^1, a^1, \dots, k^{t-1}, a^{t-1}, k^t)$$
$$\hat{h}^t = \left(\hat{k}^0, \hat{a}^0, \hat{k}^1, \hat{a}^1, \dots, \hat{k}^{t-1}, \hat{a}^{t-1}, \hat{k}^t\right).$$
- A ***Markov-perfect equilibrium*** (***MPE***) is a strategy profile satisfying
  - Perfection, i.e. at each history everyone is best responding, given the others' strategies;
  - Each player plays a Markov strategy.
- Example. Repeated prisoner's dilemma.
  - The state space has only one element, $|K| = 1$.
  - Hence, players must follow the same strategy at each period
  - Therefore, the only MPE is (D, D) every period.
  - In this example, MPE is not a really useful solution concept.
  - If we introduce states that perturb the stage game payoffs by some infinitesimal amount, then one can condition strategies on history. But this makes MPE lose its bite.

  - Examples of games where MPE applies more naturally.
    - Resource extraction: how much you extract the resources depends only on how much resources are left
    - Bequest games: behavior is forward looking
- **Theorem.** Markov perfect equilibrium exists in stochastic game with finite number of states and actions.
  - *Proof.* Construct a "Markov strategic form", i.e. a normal form game where each agent-

state pair (i,k) in the original stochastic game corresponds to a player in the new game, and where the expected payoffs are inherited from the original game. Then we know that a NE exists in the new game.
Any NE in the new game corresponds to a MPE in the original game: The NEs in the new game is Markovian, because each player in the new game is an agent-state pair. The NEs are perfect, because given any state, a player is best responding to all other players, and also best responding to herself in other states. Therefore, the theorem is established.

- Example. Suppose there is no uncertainty, so that $q(k^{t+1}|k^t, a^t) \in \{0,1\}$, and that the state space is continuous. Only one player plays in each period. Define
$$k^{t+1} = f_{t+1}(a^t)$$
(if k^{t+1} depends on k^t, then just redefine a^t to get rid of the dependence). Let $g_t(a^t, k^t)$ denote the time-t payoff, where $a^t, k^t \in \mathbb{R}$.
  - **Result.** If $\frac{\partial^2 g_t}{\partial k^t \partial a^t} \geq 0$ (≤ 0), then in an MPE, $a^t(k^t)$ is non-decreasing (non-increasing).
    - Note that $\frac{\partial^2 g_t}{\partial k^t \partial a^t} \geq 0$ is just the single-crossing (Spence-Mirrlees) condition: as k^t goes up, marginal utility from a^t goes up.
    - *Proof.* Suppose two states k and $\tilde{k}$ correspond to MPE actions a and $\tilde{a}$. Optimality implies
$$g_i^t(a,k) + \underbrace{v_i\big(f_{t+1}(a)\big)}_{\text{continuation value}} \geq g_i^t(\tilde{a},k) + v_i\big(f_{t+1}(\tilde{a})\big)$$
$$g_i^t\big(\tilde{a},\tilde{k}\big) + v_i\big(f_{t+1}(\tilde{a})\big) \geq g_i^t\big(a,\tilde{k}\big) + v_i\big(f_{t+1}(a)\big)$$
Add the two inequalities:
$$g_i^t(a,k) + g_i^t\big(\tilde{a},\tilde{k}\big) - g_i^t\big(a,\tilde{k}\big) - g_i^t(\tilde{a},k) \geq 0$$
$$\int_a^{\tilde{a}} \int_k^{\tilde{k}} \frac{\partial^2 g_i^t}{\partial x \partial y}\, dy\, dx \geq 0$$
If the cross-partial is non-negative, then either $\big(\tilde{a} \geq a \,\&\, \tilde{k} \geq k\big)$ or $\big(\tilde{a} \leq a \,\&\, \tilde{k} \leq k\big)$. Therefore, $a^t(k)$ is non-decreasing.

- What if there is no explicit state variable?
  - If you want to introduce new states, they must be payoff-relevant.
  - Let H^t be a partition of the history space $A^0 \times \cdots \times A^{t-1}$ at time t. So H^t tells you what cell each history is in. $\{H^t(\cdot)\}_{t=0}^T$ is ***sufficient*** if
$$\forall t, h^t, \tilde{h}^t : H^t(h^t) = H^t\big(\tilde{h}^t\big)$$
the subgames starting at t are strategically equivalent; that is,
    - Action spaces are identical:
$$\forall i, \tau \geq 0, a^t, \dots, a^{t+\tau-1} : A_i^{t+\tau}(h^t, a^t, \dots, a^{t+\tau-1}) = A_i^{t+\tau}\big(\tilde{h}^t, a^t, \dots, a^{t+\tau-1}\big)$$
    - The players' vNM utility functions conditional on h^t, $\tilde{h}^t$ represent the same preferences:
$$\exists \lambda_i(\cdot,\cdot) > 0, \mu_i(\cdot,\cdot,\cdot), \forall f^t \equiv (a^t, a^{t+1}, \dots, a^T) :$$
$$u_i(h^t, f^t) = \lambda_i\big(h^t, \tilde{h}^t\big) u_i\big(\tilde{h}^t, f^t\big) + \mu\big(h^t, \tilde{h}^t, f^t\big)$$
  - ***Payoff-relevant history*** is the minimal (or coarsest) sufficient partition.

- Then the payoff-relevant history can be used as the state variable to allow for MPE in a game.
- This implies that in infinite horizon games, you may want to include time in the histories.

- *Definition.* A ***Markov perfect equilibrium*** is a strategy profile that satisfies perfection and where the payoff-relevant history

$$H^t(h^t) = H^t(\tilde{h}^t) \Rightarrow \forall i : \sigma_i^t(h^t) = \sigma_i^t(\tilde{h}^t)$$

Application of MPE

- Extraction of common resource
  - Setup
    - $k^t \geq 0$ is the stock of resource at time t
    - At each t, players 1 and 2 simultaneously choose an amount $a_1^t, a_2^t \geq 0$ to extract
    - If $k^t \geq a_1^t + a_2^t$, each player i gets instantaneous payoff $g_i(a_i^t)$, and $k^{t+1} = f(k^t - a_1^t - a_2^t)$
    - If $k^t < a_1^t + a_2^t$, instantaneous payoff is $g_i(k^t/2)$, and $k^{t+1} = f(0) = 0$.
    - Assume $g_i(\cdot) \in \mathcal{C}^1$, $g_i'(\cdot) > 0$ and $g_i''(\cdot) < 0$, with $\lim_{x\to 0} g_i'(x) = \infty$
    - Assume $f \in \mathcal{C}^1$, $f' > 0$, $f'' < 0$, $f'(0) > 1/\delta$, $\lim_{x\to\infty} f'(x) < 1$
  - Goal: find MPE where strategies are continuously differentiable
    - Let $\psi(k) = k - s_1(k) - s_2(k)$. This is the remaining stock at the end of period
    - FOC from Bellman equation:
$$g_i'(s_i(k)) = \delta g_i'\left(s_i\left(f\big(\psi(k)\big)\right)\right) f'\big(\psi(k)\big)\left[1 - s_j'\left(f\big(\psi(k)\big)\right)\right]$$
      - Marginal utility next period
$$\frac{\partial g_i\left(s_i\left(f\big(\psi(k)\big)\right)\right)}{\partial \psi(k)} = g_i'\left(s_i\left(f\big(\psi(k)\big)\right)\right) s_i'\left(f\big(\psi(k)\big)\right) f'\big(\psi(k)\big) \quad (1)$$
      - Change to the leftover at the end of next period
$$f'\big(\psi(k)\big)\left[1 - s_i'\left(f\big(\psi(k)\big)\right) - s_j'\left(f\big(\psi(k)\big)\right)\right]$$
      - Change to the marginal utility from the leftover
$$g_i'\left(s_i\left(f\big(\psi(k)\big)\right)\right) f'\big(\psi(k)\big)[1 - s_i'(\cdot) - s_j'(\cdot)] \quad (2)$$
      - Adding (1) and (2), we get the FOC from Bellman
    - The FOC implies that $\psi'(k) > 0$, for otherwise, there exists $k, \tilde{k}$ with $k > \tilde{k}$ such that $s_i(k) = s_i(\tilde{k}) \;\Rightarrow\; \psi(k) > \psi(\tilde{k})$, which is a contradiction. Then,
$$\psi'(k) > 0 \;\Rightarrow\; (f \circ \psi)'(k) > 0$$
which in turn implies that the stock converges to the maximum.

Application of MPE (cont'd)

- Extraction of common resource (cont'd)
  - Solving for the FOC of the Bellman equation gives
$$g_i'(s_i(k)) = \delta g_i'\left(s_i\left(f(\psi(k))\right)\right)f'(\psi(k))\left[1 - s_j'\left(f(\psi(k))\right)\right]$$
where $\psi(k) = k - s_i(k) - s_j(k)$. This implies that
$$\psi'(k) > 0$$
because otherwise, there exists $k, \tilde{k}$ such that $k > \tilde{k}$ and
$$\begin{aligned}\psi(k) = \psi(\tilde{k}) \;&\Rightarrow\; s_i(k) = s_i(\tilde{k}) \\ &\Rightarrow\; \psi(k) > \psi(\tilde{k})\end{aligned}$$
which is a contradiction. Then, $\psi'(k) > 0$ implies that $f(\psi(\cdot))$ is strictly increasing, and thus the stock converges to a steady state.
  - In steady state, $\hat{k} = f\left(\psi(\hat{k})\right)$. Plug this into FOC:
$$\begin{aligned}g_i'\left(s_i(\hat{k})\right) = \delta g_i'\left(s_i(\hat{k})\right)f'\left(\psi(\hat{k})\right)\left[1 - s_j'(\hat{k})\right] \;&\Rightarrow\; 1 = \delta f'\left(\psi(\hat{k})\right)\left[1 - s_j'(\hat{k})\right] \\ &\Rightarrow\; s_i'(\hat{k}) = s_j'(\hat{k})\end{aligned}$$
If the steady state is stable, we need
$$s_1'(\hat{k}) = s_2'(\hat{k}) > 0 \;\Rightarrow\; 0 < 1 - s_j'(\hat{k}) < 1 \;\Rightarrow\; 1 < \delta f'\left(\psi(\hat{k})\right)$$
    - The Golden rule for k^*
$$\delta f'(\psi(k^*)) = 1$$
Since $f'' < 0$,
$$\psi(k^*) > \psi(\hat{k}) \;\Rightarrow\; k^* > \hat{k}$$
Thus, we have an intuitive result that this MPE results in an under-accumulation of capital stock (sort of like in the Cournot situation).

Social Choice Theory

- Suppose we have individuals $1, \dots, n$ with rational preferences $\succsim_1, \dots, \succsim_n$. Can we get a rational $\succsim_s$ that "sensibly" aggregates the individual preferences $\succsim_1, \dots, \succsim_n$?
  - Here we assume we have access to the true individual preferences (if we don't have access to the true preferences, then we'd be doing mechanism design theory)
  - Also, we use only ordinal information of the individual preferences (if want to use cardinal information, we'd be doing cooperative game theory)
- Formal setup of the problem
  - X is the set of social outcomes/alternatives
  - $I = \{1, \dots, n\}$ is the set of individuals
  - $\mathcal{R}$ is the set of all rational weak preference orderings on X
    - E.g. suppose $X = \{x, y\}$. Then $\mathcal{R} = \{x \succ y, x \sim y, y \succ x\}$.
- *Definition.* A ***social welfare functional*** (***SWF***) is a function $F : \mathcal{A} \to \mathcal{R}$ that assigns a social preference relation $\succsim_s \in \mathcal{R}$ to any profile of individual preference orderings $\succsim = (\succsim_1, \dots, \succsim_n) \in \mathcal{A} \subset \mathcal{R}^n$
- Some good properties for F
  - ***Universal Domain*** (***UD***): $\mathcal{A} = \mathcal{R}^n$
  - ***Symmetry/Anonymity*** (***S***): Let $\pi : \{1, \dots, n\} \to \{1, \dots, n\}$ be one-to-one. Then,
    $$F(\succsim_1, \dots, \succsim_n) = F\big(\succsim_{\pi(1)}, \dots, \succsim_{\pi(n)}\big), \qquad \forall \succsim \in \mathcal{A}$$
    - This is saying that re-ordering the <u>individuals</u> will not change the social preference. In other words, $(\succsim_1, \dots, \succsim_n)$ does not have to be ordered.
  - ***Neutrality*** (***N***): For any $\succsim, \succsim' \in \mathcal{A}$, we have
    $$\text{if} \quad x \succsim_i y \iff y \succsim'_i x, \qquad \forall i$$
    $$\text{then} \quad x \succsim_s y \iff y \succsim'_s x$$
    - This is saying that re-ordering the <u>alternatives</u> that are equally preferred to will not change the social preference
  - ***Positive Responsiveness*** (***PR***): If $x \succsim_s y$ for some profile $\succsim$, and $\succsim'$ is such that, with respect to y, x moves up in some agents' ranking and falls in no one's, then $x \succ'_s y$.
  - ***Pareto Property*** (***P***): If $x \succ_i y$ for all i, then $x \succ_s y$.
  - ***Independence of Irrelevant Alternatives*** (***IIA***): For any pair $(x, y) \in X^2$ and any profiles $\succsim, \succsim' \in \mathcal{A}$,
    $$\text{if} \quad x \succsim_i y \iff x \succsim'_i y \quad \text{and} \quad y \succsim_i x \iff y \succsim'_i x, \qquad \forall i$$
    $$\text{then} \quad x \succsim_s y \iff x \succsim'_s y \quad \text{and} \quad y \succsim_s x \iff y \succsim'_s x$$
    - The social ranking of x and y depends only on the individual rankings of x and y

- 2 alternatives: $X = \{x, y\}$. Let

$$\alpha_i = \begin{cases} +1 & \text{if } x \succ_i y \\ 0 & \text{if } x \sim_i y \\ -1 & \text{if } x \prec_i y \end{cases}$$

Preference profile is $(\alpha_1, \ldots, \alpha_n) \in \{-1,0,+1\}^n = \mathcal{A} = \mathcal{R}^n$

  - A ***simple majority SWF*** is

$$F_M : \mathcal{A} \to \{-1,0,+1\}$$

  where

$$F_M(\alpha_1, \ldots, \alpha_n) = \operatorname{sgn} \sum_{i=1}^{n} \alpha_i$$

    - This rule satisfies UD, S, N, PR.

  - **May's Theorem.** A SWF satisfies UD, S, N, PR if and only if it is the simple majority SWF (when X contains 2 outcomes).

    - *Proof.* Symmetry implies that $F_M(\alpha_1, \ldots, \alpha_n)$ can only depend on $\#(+1)$, $\#(0)$, and $\#(-1)$. Let $n^+(\succsim) \coloneqq \#(+1)$ and $n^-(\succsim) \coloneqq \#(-1)$. Let $\succsim'$ be the opposite of $\succsim$. It follows that, if $n^+(\succsim) = n^-(\succsim)$, then

$$n^+(\succsim') = n^{-1}(\succsim') \;\Rightarrow\; \succsim_s = \succsim_s'.$$

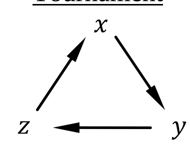
    By N, $x \sim_s y$. By PR,

$$n^+(\succsim) > n^-(\succsim) \;\Rightarrow\; x \succ_s y.$$

- 3 alternatives: $X = \{x, y, z\}$.

Person 1	Person 2	Person 3
x	y	z
y	z	x
z	x	y

Tournament



  - Can't resort to pairwise majority for transitive SWF → the preference profile violates UD. No "Condorcet winner".

- **Arrow's Impossibility Theorem.** Suppose $|X| \geq 3$. If the SWF F satisfies UD, P, and IIA, then F is dictatorial, i.e.

$$\exists i \in I, \forall x, y \in X, \forall (\succsim_1, \ldots, \succsim_n) \in \mathcal{R}^n : x \succ_i y \;\Rightarrow\; x \succ_s y.$$

Arrow's Impossibility Theorem

- Let the following conditions be satisfied:
  1. $|X| \geq 3$;
  2. A SWF F satisfies UD, P, IIA;

  then F is ***dictatorial***, that is,
$$\exists i \in I, \forall x, y \in X, \forall(\succsim_1, \dots, \succsim_n) \in \mathcal{R}^n : x \succ_i y \Rightarrow x \succ_s y$$

- *Proof.*
  - *Definition.* Suppose for some $x, y \in X$, whenever $x \succ_i y\ \forall i \in S$, and $y \succ_i x\ \forall i \notin S$, then S is ***decisive for x over y***.
  - *Definition.* Suppose for all pairs $(x, y) \in X^2$, whenever $x \succ_i y\ \forall i \in S$, and $y \succ_i x\ \forall i \notin S$, then S is ***decisive***.
  - The proof will proceed in three steps:
    1. If $S \subset I$ is decisive for x over y for some $x, y \in X$, then S is decisive.
    2. There exists an $h \in I$ such that $\{h\}$ is decisive.
    3. Remainder of the proof.
  - Step 1. We need only show that S is decisive for x over z and z over y for all $z \neq x, y$. Suppose $x \succ_i y \succ_i z$ for all $i \in S$, and $y \succ_i z \succ_i x$ for all $i \in I \setminus S$. Then, we have $x \succ_s y$ because S is decisive for x over y. Also, we have $y \succ_s z$ by the Pareto property. Thus, the two implies $x \succ_s z$. By IIA, S is decisive for x over z. Symmetrically, we must have S is decisive for z over y. [This is saying, if a group is "important" sometimes, it is "important" all the time.]
  - Step 2. There are two sub-steps:

    a) If S, T are decisive, then $S \cap T$ is decisive.
    Let $x, y, z \in X$ and suppose
$$\begin{aligned} z \succ_i y \succ_i x, &\quad \forall i \in S \setminus (S \cap T) \\ x \succ_i z \succ_i y, &\quad \forall i \in S \cap T \\ y \succ_i x \succ_i z, &\quad \forall i \in T \setminus (S \cap T) \\ y \succ_i z \succ_i x, &\quad \forall i \in I \setminus (S \cup T) \end{aligned}$$
    The assumption that S is decisive implies $z \succ_s y$. T is decisive implies $x \succ_s z$. By transitivity, $x \succ_s y$. By IIA and step 1, $S \cap T$ is decisive.

    b) $\forall S \subset T$, either S or $I \setminus S$ is decisive.
    Let $x, y, z \in X$, and suppose
$$\begin{aligned} x \succ_i z \succ_i y, &\quad \forall i \in S \\ y \succ_i x \succ_i z, &\quad \forall i \in I \setminus S \end{aligned}$$
    If $x \succ_s y$, then, by IIA, S is decisive. If $y \succsim_s x$, then by Pareto, $x \succ_s z$, and by transitivity, $y \succ_s z$, and hence by IIA, $I \setminus S$ is decisive.

    c) To complete this step, we want to show that, if $S \subset I$ is decisive and $|S| \geq 2$, then there exists $S' \subset S$, where $S' \neq S$ such that S' is decisive.

- Take $h \in S$. If $S \setminus \{h\}$ is decisive, we're done.
- If $S \setminus \{h\}$ is not decisive, then $\{h\} \cup I \setminus S$ is decisive.
- Then, $S \cap (\{h\} \cup I \setminus S) = \{h\}$ is decisive.

➢ Step 3. If S is decisive, and $S \subset T$ is decisive. Note that the Pareto property implies that $\emptyset$ is not decisive. Since $S \cap (I \setminus T) = \emptyset$, then by step 2a, $I \setminus T$ is not decisive. Then by step 2b, T is decisive.
Suppose $\{h\}$ is decisive. Pick any $T \ni h$, any $x, y \in X$. We know that T is decisive. So whenever $x \succ_i y \;\; \forall i \in T$ and $y \succ_i x \;\; \forall i \in I \setminus T$, we have $x \succ_s y$.

Therefore, $\forall x, y \in X, \exists h \in I : x \succ_h y \;\; \Rightarrow \;\; x \succ_s y$.

Arrow's Impossibility Theorem (cont'd)

- Recall from last time that
  - $\exists h \in I$ such that $\{h\}$ is decisive
  - If $S \supset I$ is decisive, and $S \subset T$, then T is also decisive
  - But note that the argument presented in the proof will not work if *some* of the people do not have strict preferences.
  - Here's a fix:
    Suppose S is decisive. Pick any $T \subset I \setminus S$, and any $x, y, z \in X$. Suppose
    $$\begin{aligned} &x \succ_i z \succ_i y, && \forall i \in S \\ &x \succsim_i y \succ_i z, && \forall i \in T \\ &y \succ_i z \succ_i x, && \forall i \in I \setminus (S \cup T) \end{aligned}$$
    Because S is decisive, $z \succ_s y$.
    Because $S \cup T$ is decisive, $x \succ_s z$.
    By transitivity, $x \succ_s y$.
  - This completes the discussion of Arrow's impossibility theorem.

Social Choice Function

- *Definition.* A ***social choice function*** (***SCF***) is a function $f : \mathcal{A} \to X$ that assigns an alternative $x \in X$ to all profiles of preferences $\succsim = (\succsim_1, \dots, \succsim_n)$ in the domain $\mathcal{A} \subseteq \mathcal{R}^n$.
  - ***Universal domain*** (***UD***): $\mathcal{A} = \mathcal{R}^n$
  - ***Pareto property*** (***P***): $\forall x, y \in X$ and $\forall \succsim \in \mathcal{A} : (\forall i : x \succ_i y) \;\Rightarrow\; f(\succsim) \neq y$
  - ***Monotonicity*** (***M***): Suppose $f(\succsim) = x$. If $\forall i \in I$ and $\forall y \neq x$, another preference profile $\succsim'$ is such that $x \succsim_i y \;\Rightarrow\; x \succsim'_i y$, then $f(\succsim') = x$
    - This replaces the IIA condition in the case of SWF.
- **Theorem.** Suppose $|X| \geq 3$. If the SCF f satisfies UD, P, and M, then f is ***dictatorial***, i.e.

$$\exists i \in I, \forall (\succsim_1, \dots, \succsim_n) \in \mathcal{R}^n : f(\succsim_1, \dots, \succsim_n) \in \operatorname*{argmax}_{x \in X} \{\succsim_i\}$$

- *Definition.* A preference profile $\succsim = (\succsim_1, \dots, \succsim_n) \in \mathcal{R}^n$ is ***single peaked*** if there exists a linear order $\geq$ on X such that $\forall i \in I, \exists x_i$ such that
  a) $x_i \succ_i y, \forall y \in X \setminus \{x\}$, and
  b) $(x_i \geq z > y \;\vee\; y < z \leq x_i) \;\Rightarrow\; (z \succ_i y), \quad \forall z, y \in X$
  - *Definition.* A ***linear order*** is a binary relation that is
    - Reflexive: $\forall x \in X : x \geq x$
    - Transitive: $\forall x, y, z \in X : (x \geq y \;\wedge\; y \geq z) \;\Rightarrow\; x \geq z$
    - Total: $\forall x \neq y \leq x : (x \geq y \;\vee\; y \geq x) \;\wedge\; \neg(x \geq y \;\wedge\; y \geq x)$
- *Definition.* Let $\succsim_M$ be the social preferences generated by ***pairwise majority***, i.e.

$$x \succsim_M y \;\Leftrightarrow\; \#\{i \in I | x \succ_i y\} \geq \#\{i \in I | y \succ_i x\}$$

- *Definition.* Individual $m \in I$ is the ***median agent/voter*** for the single peaked preference profile $\succsim \in \mathcal{R}^n$ if

$$\#\{i \in I | x_i \geq m\} \geq \frac{n}{2} \quad \text{and} \quad \#\{i \in I | x_i \leq m_m\} \geq \frac{n}{2}$$

- **Proposition.** If preference profile $\succsim \in \mathcal{R}^n$ is single peaked, then $x_m \succsim_M y$ for all $y \in X$. In other words, a Condorcet winner exists and coincides with x_m.
  - *Proof.* Suppose, to the contrary, $\exists y > x_m$ such that $y \succ_M x_m$. Then, for any $i < m$, $x_i < m < y \;\Rightarrow\; x_i \succ_i x_m \succ_i y$, and thus there are at least $\frac{n}{2}$ agents will vote for x_m over y. But this cannot happen in a pairwise majority voting.
  - This proposition only says that a Condorcet winner exists, but it does not guarantee that the social preference generated is rational.
- **Proposition.** Let P^n be the set of strict and rational preference profiles. If n is odd, and $\succ \in P^n$ is single peaked, then the social preferences generated by pairwise majority rule $\succsim_M$ are complete and transitive.
  - *Proof.* Completeness is trivial. Since the number of agents are odd, social preference is

also strict. Assume $x \succ_M y$ and $y \succ_M z$. Suppose to the contrary that $z \succ_M x$. Then we have a "cycle". But this contradicts the previous proposition, because in the case of "cycle", no Condorcet winner exists. Therefore it must be the case that $x \succ_M z$.

- *Definition.* A preference profile $\succsim \in \mathcal{R}^n$ satisfies the ***single crossing property*** if there exists a linear order $\geq$ on X and an order of agents $\{1, \dots, n\}$ such that $\forall x, y \in X, x > y$, we have

$$\forall i, \forall j > i : x \succsim_i y \;\Rightarrow\; x \succsim_j y \quad \text{and} \quad x \succ_i y \;\Rightarrow\; x \succ_j y.$$

  We have the ***strict single crossing property*** if

$$\forall i, \forall j > i : x \succsim_i y \;\Rightarrow\; x \succ_j y$$

- **Proposition (Median Voter Theorem II).** If n is odd, and preferences satisfy the strict single crossing condition, then

$$\forall x, y \in X : x \succsim_M y \;\Leftrightarrow\; x \succsim_m y$$

  where $m = \frac{1}{2}(n+1)$.

Median Voter Theorems

- Recall from last class:
  - **Median Voter Theorem I (MVT I).** If $\succsim \in \mathcal{R}^n$ is single peaked, then the median agent's peak x_m is a Condorcet winner. Furthermore, if n is odd, $\succsim_M$ is complete and transitive.
  - **Median Voter Theorem II (MVT II).** If n is odd, and $\succsim \in \mathcal{R}^n$ satisfies the strict single crossing property, then $\succsim_M = \succsim_m$.
    - But notice that $SSC \nRightarrow single\ peakedness$
- *Definition.* Let $X \subseteq \mathbb{R}^2$ and $u : X \times I \to \mathbb{R}$ with $u_2 > 0$. Then, the ***Spence-Mirrlees condition*** requires that $\frac{u_1(x,y,i)}{u_2(x,y,i)}$ be increasing in i for all $(x, y) \in \text{int}(X)$.
  - Note that implicit here is that *there exists an ordering on I* such that u_1/u_2 is increasing in i
  - The usual single crossing condition with cross-partial derivative being positive comes from the fact that u is quasi-linear in y, so that u_2 is constant. And so the cross partial $u_{13} \geq 0$ is the same as the Spence-Mirrlees condition.

Social Welfare and Cooperative Game Theory

- *Definition.* A ***utility possibility set*** is
$$U = \{(u_1, \dots, u_n) \in \mathbb{R}^n | u_1 \leq u_1(x), \dots, u_n \leq u_n(x) \text{ for some } x \in X\}$$
  - *Definition.* The ***Pareto frontier*** of the utility possibility set is
$$\{(u_1, \dots, u_n) \in \mathbb{R}^n | \nexists u' = (u'_1, \dots, u'_n) \in U : u'_i \geq u_i \ \forall i \text{ and } u'_i > u_i \text{ for some } i\}$$
- *Definition.* A ***social welfare function*** (***SWF***) is a function $W : \mathbb{R}^n \to \mathbb{R}$ that aggregates preferences: $W(u_1, \dots, u_n)$.
  - A policy maker might face a problem
$$\max_{(u_1,\dots,u_n)\in U} W(u_1, \dots, u_n)$$
This is a first-best problem. Sometimes, not all of U is available, then we have a second-best problem, where there are restrictions imposed on the set U.
- Properties on SWF:
  1. **Non-paternalism** (implied by setup): $W(\vec{u}) = W(\vec{u}')$ if $u_i = u'_i$ for all i
  2. **Pareto**: $W(\vec{u}) \geq W(\vec{u}')$ if $u_i \geq u'_i \ \forall i$, and $W(\vec{u}) > W(\vec{u})$ if $u_i > u'_i \ \forall i$
     - **Strict Pareto**: $W(\vec{u}) > W(\vec{u}')$ whenever $u_i \geq u'_i \ \forall i$ and $u_i > u'_i$ for some i
  3. **Symmetry**: $W(\vec{u}) = W(\vec{u}')$ if $\vec{u}'$ is a permutation of $\vec{u}$.
     - A function π is a ***permutation*** over $\{1, \dots, n\}$ if it is one-to-one.
     - This assumes that everyone's utility is measured on the same scale.
     - Symmetry also implies that the marginal rate of substitution (MRS) at every $\vec{u}$ where $u_i = u_j \ \forall i, j$ are all 1
  4. **Concavity (inequality aversion)**: if $W(\vec{u}) = W(\vec{u}')$, then
$$W(p\vec{u} + (1-p)\vec{u}') \geq W(\vec{u}), \qquad \forall p \in (0,1)$$
and the inequality is strict when $\vec{u} \neq \vec{u}'$.
     - If U is convex and symmetric, everyone gets the same utility at the maximum.

Axiomatic Bargaining

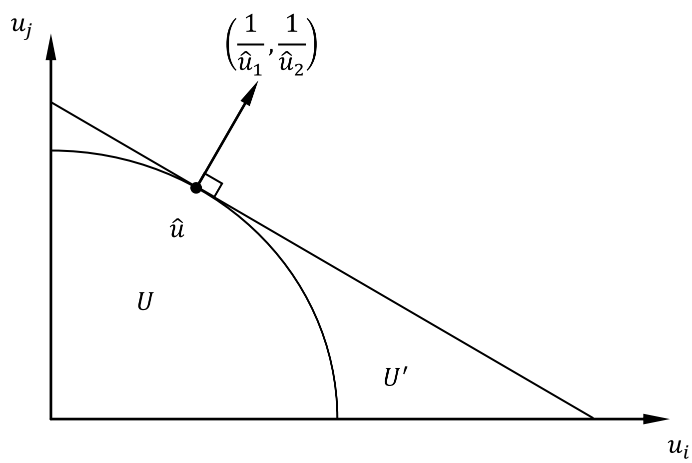
- Let $U \subseteq \mathbb{R}^n$ be the utility possibility set (UPS)
- $u^* \in U$ is the status quo
- *Definition.* A ***bargaining solution*** is a rule assigning a solution vector $f(U, u^*) \in U$ to every bargaining problem (U, u^*).
- Desirable properties of a bargaining solution
  - ***Independence of utility origins*** **(*IUO*)**:
    $$\forall \alpha \in \mathbb{R}^n, \forall i : f_i(U', u^* + \alpha) = f_i(U, u^*) + \alpha_i$$
    whenever $U' = \{(u_1 + \alpha_1, \ldots, u_n + \alpha_n) | u \in U\}$.
    - If the solution satisfies IUO, then we can normalize $u^* = \vec{0}$, since $f(U, u^*) = f\left(U - \{u^*\}, \vec{0}\right)$.
    - From now on, we'll assume IUO and write $f(U) \equiv f(U, 0)$.
  - ***Independence of utility units*** **(*IUU*)**:
    $$\forall \beta \in \mathbb{R}^n_{++}, \forall i : f_i(U') = \beta_i f_i(U)$$
    whenever $U' = \{(\beta_1 u_1, \ldots, \beta_n u_n) | u \in U\}$.
  - **(*Weak*) *Pareto property*** **(*P*)**:
    $$\forall U, \nexists u \in U, \forall i : u_i > f_i(U).$$
  - ***Symmetry*** **(*S*)**: For all symmetric U, and all i, j, $f_i(U) = f_j(U)$.
    - Here U is symmetric if $u \in U \iff$ any permutation of u is in U
  - ***Independence of Irrelevant Alternatives*** **(*IIA*)**:
    $$(U' \subset U \;\land\; f(U) \in U') \;\Rightarrow\; f(U') = f(U).$$
- Examples.
  - Egalitarian: vector in the frontier of U where all entries are equal
    - This rule satisfies S, P, IIA, but not IUU (because β_i can be different across i)
  - Utilitarian: maximize $\sum_i u_i$ (assume U is strictly convex so that solution is unique)
    - This rule satisfies S, P, IIA, but not IUU (because changing β amounts to changing the weights on people's utility, so the outcome is not invariant).
  - Nash Bargaining Solution: $\max \prod_i u_i$
    - Satisfies…
      - S, because of Cauchy-Schwartz inequality
      - P, because of maximizing
      - IIA, since maximization selects a unique argmax (assuming weak convexity of U)
      - IUU, because maximizing a product is like maximizing a sum of logs, and scaling by β is like adding a constant

- **Proposition.** The Nash bargaining solution is the <u>only</u> one satisfying IUO, IUU, P, S, and IIA.
  - *Proof.* Let f^N be the Nash bargaining solution. Suppose f satisfies all desired properties. Given U, let $\hat{u} = f^N(U)$, and let
$$U' = \left\{u \in \mathbb{R}^n \middle| \sum_i \frac{u_i}{\hat{u}_i} \le n\right\}, \qquad U'' = \left\{u \in \mathbb{R}^n \middle| \sum_i u_i \le n\right\}.$$
    We must have $U \subset U'$.
    - Note that $\sum_i \ln u_i$ is concave with gradient $\left(\frac{1}{\hat{u}_1}, \dots, \frac{1}{\hat{u}_n}\right)$ at $\hat{u}$, and reaches maximum at $\hat{u}$.
    - Note that $\left(\frac{1}{\hat{u}_1}, \dots, \frac{1}{\hat{u}_n}\right)$ is just the normal vector of $\sum_i \frac{u_i}{\hat{u}_i} = n$.

    Since U'' is symmetric, then by P and S, $f(U'') = (1, \dots, 1)$.

    By IUU, $f(U') = (\hat{u}, \dots, \hat{u}) = \hat{u}$

    By IIA, $f(U) = \hat{u} = f^N(U)$



- What if $n > 2$ and partial cooperation is possible?
  - Assume ***transferrable utility*** (***TU***), so that UPS is $\sum_i u_i \le k$
  - Let $v(S)$ be the total available utility if $S \subset I$ cooperates
    - Here $v(S)$ is called the "characteristic function" or the "worth of S"
  - Assume $v(S) \le v(I)$ for all $S \subset I$.
- The ***cooperative solution*** is a rule assigning utility allocation $f(v) \in \mathbb{R}^n$ to every game $v(\cdot)$ such that $\sum_i f_i(v) \le v(I)$.
- Properties
  - Independence of utility origins and of common changes of utility units (IUU).
$$\left(\forall S \subset I, \forall \beta > 0 : v(S) = \beta v'(S) + \sum_{i \in S} \alpha_i\right) \Rightarrow f(v) = \beta f(v') + (\alpha_1, \dots, \alpha_n)$$

- Pareto (P).

$$\sum_{i \in I} f_i(v) = v(I)$$

- Symmetry (S). If $v'(S) = v\big(\pi(S)\big)$ for all $S \subset I$ and permutation π, then

$$f_i(v') = f_{\pi(i)}(v), \qquad \forall i$$

- Dummy Axiom (D).

$$\forall v, \forall i, : v(S \cup \{i\}) = v(S), \qquad \forall S \subset I, \qquad f_i(v) = 0$$

- Linearity (L): if $u(S) = w(S) + v(S)$ for all S, then $f(u) = f(w) + f(v)$ for all $i \in I$

Axiomatic Bargaining (cont'd)

- Recall the desired properties of the *cooperative bargaining solution* with transferrable utility
  - Independence of utility origins & common changes of utility units (IUU)
  - Pareto (P)
  - Symmetry (S)
  - Dummy axiom (D)
  - Linearity (L)
- *Definition*. Let $g_{v,\pi(i)} = v(\{h : \pi(h) \le \pi(i)\}) - v(\{h : \pi(h) < \pi(i)\})$, where $\pi(\cdot)$ is a permutation. The ***Shapley value*** is

$$f_i^{\text{Shapley}}(v) = \frac{1}{n!}\sum_{\pi} g_{v,\pi}(i).$$

- Example. Suppose the cost of visiting three schools $\{1,2,3\}$ is described as follows

visiting	1	2	3	1&2	2&3	1&3	1&2&3
Cost	800	800	800	1000	1000	1400	1600

Thus, the Shapley value for 1 is

$$\left.\begin{matrix} 123 & 800 \\ 132 & 800 \\ 213 & 200 \\ 231 & 600 \\ 312 & 600 \\ 321 & 600 \end{matrix}\right\} \Rightarrow f_1^{\text{Shapley}} = \frac{3600}{6} = 600$$

Similarly, verify that $f_2^{\text{Shapley}} = 400$ and $f_3^{\text{Shapley}} = 600$.

- **Proposition.** The Shapley value is the only solution satisfying IUU, P, S, D, and L.
  - *Proof*. Consider the "T-unanimity game" for all $T \subseteq I, T \neq \emptyset$:

$$v_T(S) = \begin{cases} 1 & \text{if } T \subseteq S \\ 0 & \text{otherwise} \end{cases}$$

By D, S, P, we must have

$$f_i(v_T) = \begin{cases} \frac{1}{|T|} & \text{if } i \in T \\ 0 & \text{if } i \notin T \end{cases}$$

By IUU and L, it suffices to show that there is a unique way to write any v as a linear combination of v_T, i.e. that the set of T-unanimity games is linearly independent. Suppose, to the contrary, that the set of T-unanimity games were linearly dependent. Then,

$$\exists\{\beta_T\}_{T\subseteq I, T\neq\emptyset} \neq (0,\dots,0), \forall S : \sum_{T\subseteq I, T\neq\emptyset} \beta_T v_T(S) = 0.$$

Let T_0 be such that $\beta_T = 0$ for all $T \subsetneq T_0$ and $\beta_{T_0} \neq 0$. But then

$$\sum_{T\subseteq I, T\neq\emptyset} \beta_T v_T(T_0) = \underbrace{\sum_{T\subsetneq T_0} \beta_T v_T(T_0)}_{=0} + \beta_{T_0} \underbrace{v_{T_0}(T_0)}_{=1} + \underbrace{\sum_{T\not\subseteq T_0} \beta_T v_T(T_0)}_{=0} = \beta_{T_0} \neq 0.$$

This is a contradiction. So it must be the case that the set of T-unanimity games is linearly independent. But how does this relate to the Shapley value? Consider

$$v(\{i\}) = 800,$$
$$v(\{1,2\}) = v(\{2,3)\} = 1000, \qquad v(\{1,3\}) = 1400,$$
$$v(I) = 1600$$

Now decompose the original game with the above characteristic function v into T-unanimity games

$$v = 800v_{\{1\}} + 800v_{\{2\}} + 800v_{\{3\}} - 600v_{\{1,2\}} - 600v_{\{2,3\}} - 200v_{\{1,3\}} + 600v_I$$

This concludes the proof.

- *Definition.* A characteristic function is ***superadditive*** if
$$\forall S,T \subset I : S \cap T = \emptyset \;\Rightarrow\; v(S) + v(T) \leq v(S \cup T)$$
- *Definition.* In game $v(\cdot)$, $f \in \mathbb{R}^n$ is ***blocked*** by $S \subseteq I$ if $\sum_{i\in S} f_i < v(S)$.
- *Definition.* Outcome f is in the ***core*** if $\nexists S$ that blocks f.

Mechanism Design

- Setup
  - X is a set of alternatives
  - $I = \{1, \dots, n\}$ is the set of players with preferences $\succsim_i (\theta_i)$ on X. Assume the preferences can be represented by vNM utility function $u_i : X \times \Theta_i \to \mathbb{R}$
  - $\Theta = (\Theta_i)_{i \in I}$ is the set of possible states of the world, $\theta = (\theta_1, \dots, \theta_n)$, and the states of the world determine the preference profile $\succsim \in \mathcal{R}^n$
  - Social planner tries to implement a social choice function $f : \Theta \to X$.
    - There are different measures of efficiency
      - Ex-ante: before anybody observes anything
      - Interim: players observes their types, but before playing the game
      - Ex-post: after the game is played, and all player types are revealed
- *Definition.* A ***mechanism*** is a game $\big(S_1, \dots, S_n, g(\cdot)\big)$ where $g : S_1 \times \cdots \times S_n \to X$.
  - We typically assume that the planner can commit to g.
- *Definition.* A mechanism **(*fully*)** ***implements*** the social choice function $f(\cdot)$ if the (unique) equilibrium outcome of the mechanism in state θ is $f(\theta)$, i.e. $g\big(s_1^*(\theta), \dots, s_n^*(\theta)\big) = f(\theta)$.
  - Equilibrium here can be different things, e.g.
    - Dominant strategies equilibrium
    - Bayesian Nash equilibrium

- Example. Suppose there is a public project, with cost $c > 0$, the set of alternatives

$$X = \left\{ (y, t_1, \dots, t_n) \middle| y \in \{0,1\}, \qquad t_i \in \mathbb{R}, \qquad \sum_i t_i \geq cy \right\}$$

and utilities

$$u_i(x, \theta_i) = \theta_i y - t_i$$

the states of the world

$$\Theta = \{(\theta, \bar{\theta}, \dots, \bar{\theta}) \big| \theta \in [0, \infty), \qquad \bar{\theta} \text{ constant}, \qquad \bar{\theta} < (n-1)c\, \}$$

  - Consider the SCF

$$y(\theta) = 1 \iff \sum_i \theta_i \geq c, \qquad t_i(\theta) = \frac{1}{n} c y(\theta)$$

This is not implementable if

$$\left(\theta < \frac{c}{n} \;\&\; (n-1)\bar{\theta} \in \left(\frac{c(n-1)}{n}, c\right)\right) \vee \left(\theta > \frac{c}{n} \;\&\; (n-1)\bar{\theta} < \frac{c(n-1)}{n}\right)$$

In the former, the player is going to understate his valuation to avoid , and in the latter, he will overstate to get the project built.

- Example. First-price sealed bid auction
  - 2 agents: $i = 1,2$
  - Principal with zero valuation for the object
  - $X = \{(y_1, y_2, t_1, t_2) | y_i \in \{0,1\},\ \sum_i y_i = 1,\ t_i \in \mathbb{R},\ \sum_i t_i \geq 0\}$
  - Utilities $u_i(X, \theta_i) = \theta_i y_i - t_i$, where $\theta_i \overset{iid}{\sim} U_{[0,1]}$.
  - Is the following SCF implementable
    $$y_1(\theta) = 1 \;\Leftrightarrow\; \theta_1 \geq \theta_2, \qquad t_i(\theta) = \theta_i y_i(\theta)$$
    No, because reporting truthfully is not an equilibrium, given the other player's truth telling.
  - The following SCF is implementable
    $$y_1(\theta)1 \;\Leftrightarrow\; \theta_1 \geq \theta_2, \qquad t_i = \frac{1}{2}\theta_i y_i(\theta)$$
    Expected revenue of the auctioneer is
    $$\frac{1}{2}(\text{higher value}) = \frac{1}{2}\int_0^1 x(x) + (1-x)\left(\frac{1+x}{2}\right)dx = \frac{1}{3}$$
- Example. Second price sealed bid auction
  - This implements (in dominant strategy)
    $$y_1(\theta) = 1 \;\Leftrightarrow\; \theta_1 \geq \theta_2, \qquad t_i(\theta) = \theta_j y_i(\theta), \qquad i \neq j$$
    Expected revenue of auctioneer is also $1/3$.

- *Definition.* A ***direct revelation mechanism*** is a mechanism $\left(S_1, \dots, S_n, g(\cdot)\right)$ in which $S_i = \Theta_i$ for all i, and $g(\theta) = f(\theta)$.
- *Definition.* $f(\theta)$ is ***truthfully implementable*** (or ***incentive compatible***) if the direct revelation mechanism has an equilibrium in which
  $$\forall i \in I, \forall \theta_i \in \Theta_i : s_i^*(\theta_i) = \theta_i.$$
  In other words, truth-telling constitutes an equilibrium.
- *Definition.* Strategy $s_i : \Theta_i \to S_i$ is ***weakly dominant*** for player i if
  $$\forall s_{-i} \in S_{-i}, \forall \theta_i \in \Theta_i, \forall \tilde{s}_i \in S_i : u_i[g(s_i(\theta_i), s_{-i}), \theta_i] \geq u_i[g(\tilde{s}_i, s_{-i}), \theta_i]$$
- *Definition.* Strategy profile $\left(s_1^*(\theta_1), \dots, s_n^*(\theta_n)\right)$ is a ***dominant strategy equilibrium*** if
  $$\forall i \in I, \forall \tilde{s}_i \in S_i, \forall \theta_i \in \Theta_i, \forall s_{-i} \in S_{-i} : u_i[g(s_i^*(\theta_i), s_{-i}), \theta_i] \geq u_i[g(\tilde{s}_i, s_{-i}), \theta_i]$$

Mechanism Design (cont'd)

- *Definition.* Mechanism Γ ***implements the SCF $f(\theta)$ in dominant strategies*** if there exists a dominant strategy equilibrium $(s_1^*(\theta), \ldots, s_n^*(\theta))$ of Γ such that
$$g(s_1^*(\theta), \ldots, s_n^*(\theta)) = f(\theta), \qquad \forall \theta \in \Theta.$$
- **Theorem (Revelation Principle for dominant strategies).** Suppose the SCF $f(\theta)$ is implementable in dominant strategies. Then $f(\theta)$ is also truthfully implementable in dominant strategies; that is, there exists $\Gamma = (\Theta_1, \ldots, \Theta_n, f(\theta))$ that has an equilibrium where
$$\forall i \in I, \forall \theta_i, \tilde{\theta}_i \in \Theta_i, \forall \theta_{-i} \in \Theta_{-i} : u_i[f(\theta_i, \theta_{-i}), \theta_i] \geq u_i[f(\tilde{\theta}_i, \theta_{-i}), \theta_i].$$
    - *Proof.* We know that there exists a mechanism with equilibrium $(s_1^*(\theta), \ldots, s_n^*(\theta))$ and outcome $g(s(\theta)) = f(\theta)$ for all $\theta \in \Theta$. This is to say that
    $$\forall i \in I, \forall \tilde{s}_i \in S_i, \forall s_{-i} \in S_{-i}, \forall \theta_i \in \Theta_i : u_i[g(s_i^*(\theta_i), s_{-i}), \theta_i] \geq u_i[g(\tilde{s}_i, s_{-i}), \theta_i].$$
    In particular, this must be true for
    $$s_{-i} = s_{-i}^*(\theta_{-i}) \quad \text{and} \quad \tilde{s}_i = s_i^*(\tilde{\theta}_i).$$
    This implies that
    $$u_i[g(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})), \theta_i] \geq u_i\left[g\left(s_i^*(\tilde{\theta}_i), s_{-i}^*(\theta_{-i})\right), \theta_i\right].$$
    Note that $g(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})) = f(\theta_1, \ldots, \theta_n)$ for all $\theta \in \Theta$. Then, the above inequality becomes
    $$u_i[f(\theta_i, \theta_{-i}), \theta_i] \geq u_i[f(\tilde{\theta}_i, \theta_{-i}), \theta_i].$$
    This completes the proof.
        - Notice the importance of the commitment of the principal. If the principal were not able to commit, then the agents may not believe that $g(s(\theta)) = f(\theta)$, and so we cannot make the last substitution.
        - Implication: If $f(\theta_i, \theta_{-i}) \neq f(\tilde{\theta}_i, \theta_{-i})$, then there is a preference reversal when i's type changes from θ_i to $\tilde{\theta}_i$, i.e.
        $$u_i[f(\theta_i, \theta_{-i}), \theta_i] \geq u_i[f(\tilde{\theta}_i, \theta_{-i}), \theta_i]$$
        but
        $$u_i[f(\theta_i, \theta_{-i}), \tilde{\theta}_i] \leq u_i[f(\tilde{\theta}_i, \theta_{-i}), \tilde{\theta}_i].$$
        - Implication: $f(\theta)$ must be ***monotonic***, i.e. $\forall \theta$, if θ' is such that the lower contour set $L_i(f(\theta), \theta_i) \subseteq L_i(f(\theta), \theta')$ for all i, then $f(\theta') = f(\theta)$.

Dominant Strategy Implementation

- Recall the two implications of the revelation principle:
  - Preference reversal. For all θ_{-i},
$$u_i(f(\theta_i, \theta_{-i}), \theta_i) \geq u_i\big(f\big(\tilde{\theta}_i, \theta_{-i}\big), \theta_i\big)$$
but
$$u_i\big(f(\theta_i, \theta_{-i}), \tilde{\theta}_i\big) \leq u_i\big(f\big(\tilde{\theta}_i, \theta_{-i}\big), \tilde{\theta}_i\big).$$
  - Monotonicity of $f(\theta)$:
$$\forall \theta, \theta' : \big(\forall i : L_i(f(\theta), \theta_i) \subseteq L_i(f(\theta), \theta_i')\big) \;\Rightarrow\; \big(f(\theta') = f(\theta)\big).$$
Suppose $\theta' = (\theta_i', \theta_{-i})$. The fact that f is truthfully implementable implies that neither θ_i nor θ_i' wants to lie. Thus,
$$\theta_i \text{ truthful} \;\Rightarrow\; f(\theta_i', \theta_{-i}) \in L_i(f(\theta), \theta_i)$$
$$\theta_i' \text{ truthful} \;\Rightarrow\; f(\theta) \in L_i(f(\theta_i', \theta_{-i}), \theta_i')$$
- **Gibbard-Satterthwaite Theorem.** Suppose $|X| \geq 3$, $\mathcal{R}_i = \mathcal{P}$ for all i, and $f(\Theta) = X$. Then, $f(\theta)$ is truthfully implementable in dominant strategies if and only if it is dictatorial. That is,
$$\exists i \in I : f(\theta) \in \operatorname*{argmax}_{x \in X} u_i(x, \theta_i).$$
  - Recall that if $|X| \geq 3$, $\mathcal{R}_i = \mathcal{P}$, and f is Paretian and monotonic, then f is dictatorial.
  - **Lemma.** If f is monotonic and onto (i.e. $f(\Theta) = X$), then f is *ex post* efficient.

Dominant Strategy Implementation (cont'd)

- **Gibbard-Satterthwaite Theorem.** Suppose $|X| \geq 3$, $\mathcal{R}_i = \mathcal{P}$ for all i, and $f(\Theta) = X$. Then, $f(\theta)$ is truthfully implementable in dominant strategies if and only if it is dictatorial. That is,
$$\exists i \in I : f(\theta) \in \underset{x \in X}{\operatorname{argmax}}\, u_i(x, \theta_i).$$
  - **Lemma.** If f is monotonic and onto (i.e. $f(\Theta) = X$), then f is *ex post* efficient.
    - *Proof.* Suppose not. Then there exists some outcome $y \in X$ such that
$$u_i(y, \theta_i) > u_i(f(\theta), \theta_i), \qquad \forall i.$$
Because f is onto, there exists some θ' such that $y = f(\theta')$. This implies that
$$u_i(f(\theta'), \theta_i) > u_i(f(\theta), \theta_i), \qquad \forall i.$$
Choose $\theta'' \in \Theta$ such that
$$u_i(f(\theta'), \theta_i'') > u_i(f(\theta), \theta_i'') > u_i(z, \theta_i''), \qquad \forall z \neq f(\theta'), f(\theta).$$
By Monotonicity, $f(\theta') = f(\theta'')$. By Monotonicity again, $f(\theta'') = f(\theta)$. But this implies that $f(\theta') = f(\theta)$, which contradicts our assumption.

- ***"Groves-Clarke" Environment***. Let $X = \{(y, t_1, \ldots, t_n) | y \in Y, t_i \in \mathbb{R}, \sum_i t_i \leq 0\}$, and
$$u_i(x, \theta_i) = v_i(y, \theta_i) + t_i$$
  - Think about t_i as transfer *to* agent i, so that $t_i < 0$ means that i is paying money.
- *Definition.* $y^*(\theta)$ is ***efficient*** if and only if
$$\sum_i v_i(y^*(\theta), \theta_i) \geq \sum_i v_i(y, \theta_i), \qquad \forall y \in Y.$$
- **Theorem.** A ***Groves-Clarke mechanism*** is a SCF $f(\theta)$ with an efficient decision $y^*(\theta)$ that is implementable in dominant strategies, where $f(\theta)$ is defined as
$$y^*(\theta_i, \tilde{\theta}_{-i}) \in \underset{y \in Y}{\operatorname{argmax}} \sum_i v_i(y, \theta_i)$$
$$t_i(\theta) = \sum_{j \neq i} v_j(y^*(\theta), \theta_j) + \underbrace{h_i(\theta_{-i})}_{\substack{i \text{ cannot} \\ \text{influence}}}, \qquad \forall \theta \in \Theta$$
  - *Proof.* Suppose the Groves-Clarke mechanism does not implement $y^*(\theta)$ is dominant strategies. Then there exists $\hat{\theta}_i \neq \theta_i$ that is better for i to report in some state θ if others announce some $\tilde{\theta}_{-i}$. This implies that
$$u_i(\cdot)|_{\text{report } \hat{\theta}_i} > u_i(\cdot)|_{\text{report } \theta_i}$$
$$v_i(y^*(\hat{\theta}_i, \tilde{\theta}_{-i}), \theta_i) + \sum_{j \neq i} v_j(y^*(\hat{\theta}_i, \tilde{\theta}_{-i}), \tilde{\theta}_j) + h_i(\tilde{\theta}_{-i})$$
$$> v_i(y^*(\theta_i, \tilde{\theta}_{-i}), \theta_i) + \sum_{j \neq i} v_j(y^*(\theta_i, \tilde{\theta}_{-i}), \tilde{\theta}_j) + h_i(\tilde{\theta}_{-i})$$
By definition of $y^*(\theta)$, truth-telling is the best response.
  - The Groves-Clarke mechanism is the *only* SCF that implements the efficient outcome if the class of function v_i is large enough.

- The "Clarke" part of the mechanism is to specify that

$$h_i(\theta_-) = -\sum_{j \neq i} v_j\left(y^*_{-i}(\theta_{-i}), \theta_j\right)$$

where

$$y^*_{-i}(\theta_{-i}) \in \operatorname*{argmax}_{y \in Y} \sum_{j \neq i} v_j\left(y, \theta_j\right).$$

  - Note that the second price auction is an example of the Clarke mechanism.

Bayesian Nash Implementation

- Setup
  - State $\theta = (\theta_1, \dots, \theta_n)$ drawn from $\Theta = \Theta_1 \times \cdots \times \Theta_n$ according to probability (density) function $\phi(\theta)$.
  - A vNM utility function $u_i(x, \theta_i)$
  - θ_i is privately observed by i
  - Each i holds beliefs about θ_{-i} according to $\phi(\theta)$, and this is a common knowledge
- Want to design a mechanism $\Gamma = (I, \Theta, u_i(\theta), S, g)$, which is a game of incomplete (and asymmetric) information → natural solution concept is the Bayesian NE.
- *Definition.* A mechanism Γ ***implements SCF $f(\theta)$ in Bayesian NE*** if there exists a BNE of Γ, $\left(s_1^*(\theta), \dots, s_n^*(\theta)\right)$ such that $g\left(s_1^*(\theta), \dots, s_n^*(\theta)\right) = f(\theta)$ for all $\theta \in \Theta$.
  - Truthful implementation is when $S_i(\theta) = \Theta_i$, and $s_i^*(\theta) = \theta_i$.
- **Revelation Principle in BNE.** If SCF $f(\theta)$ is implementable in BNE, then $f(\theta)$ is also truthfully implementable in BNE; that is, there exists a mechanism Γ_D with an equilibrium where

$$\mathrm{E}_{\theta_{-i}}[u(f(\theta_i, \theta_{-i}), \theta_i)|\theta_i] \geq \mathrm{E}_{\theta_{-i}}\left[u\left(f\left(\tilde{\theta}_i, \theta_{-i}\right), \theta_i\right)\middle|\theta_i\right], \qquad \forall i \in I, \forall \theta_i, \tilde{\theta}_i \in \Theta_i.$$

- ***D'Apremont, Gerard-Varet (expected externality) mechanism***.

$$y^*(\theta) \in \operatorname*{argmax}_{y \in Y} \sum_i v_i(y, \theta_i)$$

$$t_i(\theta) = \underbrace{\mathrm{E}_{\theta_{-i}}\left[\sum_{j \neq i} v_j\left(y^*(\theta), \theta_j\right)\right]}_{\text{expected externality}} + h_i(\theta_{-i}), \qquad \forall \theta \in \Theta$$

  - $X = \{(y, t_1, \dots, t_n) | y \in Y, t_i \in \mathbb{R}, \sum_i t_i \leq 0\}$ and $u_i(x, \theta_i) = v_i(y, \theta_i) + t_i$
  - Assume that θ_i are drawn *independently*, i.e. $\phi(\theta) = \prod_i \phi_i(\theta_i)$
  - $f(\theta)$ is (*ex post*) efficient if and only if

$$\sum_i v_i(y^*(\theta), \theta_i) \geq \sum_i v_i(y, \theta_i), \qquad \forall y \in Y$$

    but also require $\sum_i t_i(\theta) = 0$.
  - Why does this mechanism work as it claims: by definition of y^*

$$\mathrm{E}_{\theta_{-i}}\left[v_i\left(y^*\left(\tilde{\theta}_i, \theta_{-i}\right), \theta_i\right) + \sum_{j \neq i} v_j\left(y^*\left(\tilde{\theta}_i, \theta_{-i}\right), \theta_j\right) + h_i(\theta_{-i})\right]$$
$$\leq \mathrm{E}_{\theta_{-i}}\left[v_i(y^*(\theta_i, \theta_{-i}), \theta_i) + \sum_{j \neq i} v_j\left(y^*(\theta_i, \theta_{-i}), \theta_j\right) + h_i(\theta_{-i})\right]$$

    Note that the expected externality depends on θ_i but not θ_{-i}. Denote that as $H_i(\theta_i)$. Then we can get *ex post* budget balance by letting

$$h_i(\theta_{-i}) = -\sum_{j \neq i} \frac{H_j(\theta_j)}{n-1}$$

which only depends on θ_j for $j \neq i$.

- Problem: may fail the "participation constraint"

Bayesian Nash Implementation (cont'd)

- Budget balance in Groves and AGV mechanisms
  - Budget balance requires that

$$\sum t_i(\theta) = 0, \qquad \forall \theta \in \Theta$$

    - Degrees of freedom for transfer schemes: $|\Theta| \cdot n$
    - Restrictions imposed by budget balance: $|\Theta|$
  - Groves mechanism:

$$\text{pay externalities on others} + h_i(\theta_{-i})$$

    - Suppose $\Theta_i = \{\theta_i^1, \dots, \theta_i^k\}$
    - If i says θ_i^1, we can normalize i's payoff using $h_i(\cdot)$.
    - So Groves mechanism imposes, for each i,

$$(|\Theta_i| - 1) \cdot |\Theta_i| = |\Theta| - |\Theta_{-i}|.$$

number of restrictions. Thus, the total restrictions for all i is

$$|\Theta| + \sum_{i=1}^{n} (|\Theta| - |\Theta_{-i}|) = (n+1)|\Theta| = \sum_{i=1}^{n} |\Theta_{-i}|$$

If $|\Theta_i| > n$ for all i, then

$$\sum_{i=1}^{n} |\Theta_{-i}| = \sum_{i=1}^{n} \frac{|\Theta|}{|\Theta_{-i}|} < |\Theta|$$

There is more restrictions than degrees of freedom.
  - AGV mechanism:

$$\text{pay expected externalities on others} + h_i(\theta_{-i})$$

    - AGV mechanism imposes, for each i, $(|\Theta_i| - 1)$.
    - This implies a total of

$$|\Theta| + \sum_{i=1}^{n} (|\Theta_i| - 1) \ll |\Theta| \cdot n, \qquad \text{if } |\Theta_i| \gg n$$

- Example (auction of one indivisible object, n bidders, with $\theta_i \overset{iid}{\sim} U_{[0,1]}$).
  - Groves mechanism: second-price auction
  - AGV mechanism:
    - The probability of $\theta_i \geq \theta_j$ for all $j \neq i$ is θ_i^{n-1}.
    - If every $j \neq i$ are below θ_i, then $\theta_j \overset{iid}{\sim} U_{[0,\theta_i]}$
    - Let θ be the highest valuation among the $n-1$ players. Let $g(\theta)$ denoted the distribution of θ. Then,

$$g(\theta) = \frac{1}{\theta_i}\left[\left(\frac{\theta}{\theta_i}\right)^{n-2}(n-1)\right]$$

- expected externality of i with type θ_i is

$$\begin{aligned}\theta_i^{n-1}\int_0^{\theta_i} g(\theta)\theta d\theta &= \theta_i^{n-1}\int_0^{\theta_i}\frac{\theta^{n-1}}{\theta_i^{n-1}}(n-1)d\theta \\ &= \theta_i^{n-1}\frac{[\theta^n]_0^{\theta_i}}{\theta_i^{n-1}}\cdot\frac{n-1}{n} \\ &= \theta_i^{n-1}\theta_i\frac{n-1}{n} \\ &= \theta_i^n\frac{n-1}{n}\end{aligned}$$

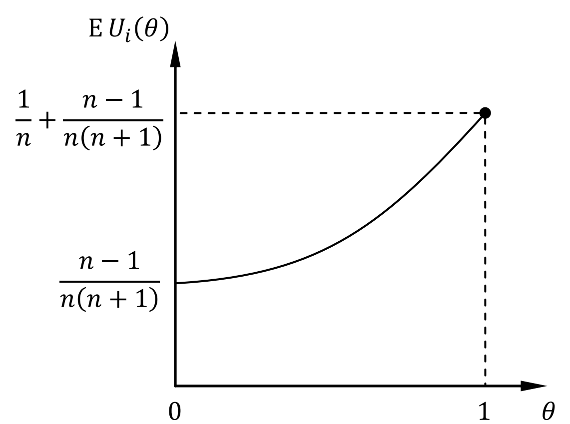
- So in the AGV mechanism the winner pays

$$\theta_i^n\frac{n-1}{n} + h_i(\theta_{-i}) = \theta_i^n\frac{n-1}{n} + \frac{1}{n-1}\sum_{j\neq i}\theta_j^n\frac{n-1}{n}$$

and this balances the budget.

Revenue Equivalence Theorem

- Recap: n bidders, $\theta_i \overset{iid}{\sim} U_{[0,1]}$
  - AGV mechanism balances budget by making each i pay
$$t_i(\theta) = -\frac{n-1}{n}\theta^n + \sum_{j\neq i}\frac{1}{n-1}\cdot\frac{n-1}{n}\theta_j^n$$
  - A principal using the AGV mechanism can implement it by charging an entrance fee, and give the object to the person with the highest valuation.
- How much can the principal charge? Note that
$$\begin{aligned}\mathrm{E}_\theta\, U_i(\theta) &= \theta\cdot\theta^{n-1} + \mathrm{E}\, t_i(\theta)\\ &= \theta^n - \frac{n-1}{n}\theta^n + \frac{n-1}{n}\mathrm{E}\big[\theta_j^n\big]\\ &= \frac{1}{n}\theta^n + \frac{n-1}{n}\int_0^1 \theta_j^n d\theta_j\\ &= \frac{1}{n}\theta^n + \frac{n-1}{n}\left[\frac{\theta_j^{n+1}}{n+1}\right]_0^1\\ &= \frac{\theta^n}{n} + \frac{n-1}{n(n+1)}\end{aligned}$$



Thus, the maximum amount that can be charged while satisfying everyone's participation constraint is $\frac{n-1}{n(n+1)}$. Consequently, the total revenue from entrance fee into AGV mechanism is $\frac{n-1}{n+1}$.

- Compare this to the expected revenue from the second price auction:
$$\begin{aligned}\mathrm{E}\, R(\theta) &= \int_0^1 n(n-1)(1-\theta)\theta^{n-2}\theta\, d\theta\\ &= n(n-1)\int_0^1(\theta^{n-1}-\theta^n)d\theta\\ &= n(n-1)\left[\frac{\theta^n}{n} - \frac{\theta^{n+1}}{n+1}\right]_0^1\\ &= n(n-1)\left(\frac{1}{n}-\frac{1}{n+1}\right)\\ &= \frac{n-1}{n+1}\end{aligned}$$

This is the same as the expected revenue in the AGV mechanism (and also first-price auction)

- **Revenue Equivalence Theorem.** Suppose we have
  - n risk neutral buyers
  - each buyer receives a private signal θ_i about value of object

- θ_i drawn <u>independently</u> from $[\underline{\theta}_i, \bar{\theta}_i]$, with density $\phi_i(\theta_i) > 0$.

Then, any two auctions where

1. the object goes to buyer with highest signal; <u>AND</u>
2. buyer i's type $\underline{\theta}_i$ has the same surplus,

yield the same expected revenue for the seller.

- Note: Conditions 1 and 2 mean that, if buyers are not symmetric (in the sense that $\phi_i(\theta_i) = \phi(\theta_i)$ for all i), then it is hard to find two auctions that satisfy both conditions.
- *Proof.* Let $\Phi_{-i}(\theta) = \prod_{j\neq i} \Phi_j(\theta) = \Pr(\theta > max_{j\neq i}\{\theta_j\})$, and $\phi_{-i}(\theta) = \Phi'_{-i}(\theta)$. Let $\hat{\theta}$ be the <u>reported</u> type and θ the <u>true</u> type.

$$\begin{aligned} \mathrm{E}\, u_i(\theta, \hat{\theta}) &= \theta \Phi_{-i}(\hat{\theta}) + \mathrm{E}\, t_i(\hat{\theta}) \\ &= \theta \Phi_{-i}(\hat{\theta}) + \bar{t}_i(\hat{\theta}). \end{aligned}$$

Truth telling requires that

$$\left.\frac{\partial\, \mathrm{E}\, u_i(\theta, \hat{\theta})}{\partial \hat{\theta}}\right|_{\hat{\theta}=\theta} = \theta \phi_{-i}(\theta) + \bar{t}'_i(\hat{\theta}) = 0$$

$$\Rightarrow \quad \bar{t}_i(\theta) = \bar{t}_i(\underline{\theta}_i) - \int_{\underline{\theta}_i}^{\theta} \theta' \phi_{-i}(\theta') d\theta'$$

This completes the proof, because the first term, $\bar{t}_i(\underline{\theta}_i)$, is given, and the second term is a constant.

- What if buyers are risk averse?
  - In second price auction, telling the truth is a dominant strategy, so risk aversion doesn't really change anything.
  - In first price auction, buyers bid under the true value. But bidding below the true value entails a risk of not winning when you should.
- What if signals are not private and independent.
  - If we have correlated common signal, then truth telling can be induced using some "crazy side bet": for buyers whose values are correlated, if their reported types do not sufficiently resemble the correlation, then they get a huge negative payoff. This way, the truth telling requirement in the proof does not have to hold anymore.

❖ **Myerson-Satterthwaite Theorem.** Suppose we have

- a bilateral trade with risk-neutral buyer and seller
- buyer's valuation $\theta_b \in [\underline{\theta}_b, \bar{\theta}_b]$
- seller's valuation $\theta_s \in [\underline{\theta}_s, \bar{\theta}_s]$
- atomless positive density functions
- $(\underline{\theta}_b, \bar{\theta}_b) \cap (\underline{\theta}_s, \bar{\theta}_s) \neq \emptyset$.

Then, there does not exists an *ex post* efficient (i.e. efficient and budget balance), Bayesian incentive compatible (i.e. truth telling) SCF satisfying participation constraints for all types.

- *Proof*. Consider the following mechanism (where $\hat{\theta}_i$ is the reported type of i):
  - If $\hat{\theta}_b \leq \hat{\theta}_s$, nothing happens
  - If $\hat{\theta}_b > \hat{\theta}_s$, then seller gets $\min\{\bar{\theta}_s, \hat{\theta}_b\}$ and buyer pays $\max\{\underline{\theta}_b, \hat{\theta}_s\}$

  In this mechanism,
  - agent with the highest value gets the good
  - incentive compatibility is satisfied
  - seller with type $\bar{\theta}_s$ gets utility $\bar{\theta}_s$ in mechanism, which is also the minimum required for participation
  - buyer with type $\underline{\theta}_b$ gets utility 0 in mechanism, which is also the minimum required for participation

  By the revenue equivalence theorem, any mechanism giving object to higher-value agent that is incentive compatible and gives $\bar{\theta}_s$ and $\underline{\theta}_b$ yields the same expected revenue. But notice that the mechanism does not satisfy budget balance: when there is transaction, i.e. $\hat{\theta}_b > \hat{\theta}_s$, buyer is paying $\hat{\theta}_s$, but seller is receiving $\hat{\theta}_b$. This means that the mechanism will run a deficit in expectation. So budget will not balance in expectation.

All-Pay Auction

- Environment: n bidders, $\theta_i \overset{iid}{\sim} U_{[0,1]}$
- Using the insight from the revenue equivalence theorem, let

$$u_i(0) = 0$$
$$u_i(\theta) = \int_0^\theta u_i'(\theta')d\theta'$$

where

$$u_i(\theta,\hat{\theta}) = \theta\hat{\theta}^{n-1} - b_i(\hat{\theta}) \;\Rightarrow\; \left.\frac{\partial u_i(\theta,\hat{\theta})}{\partial\hat{\theta}}\right|_{\hat{\theta}=\theta} = \theta(n-1)\hat{\theta}^{n-2} - b'(\hat{\theta}) = 0$$
$$\Rightarrow\; (n-1)\theta^{n-1} = b_i'(\theta)$$

Therefore, the bidding function in all pay auction is

$$b_i(\theta) = \int_0^\theta (n-1){\theta'}^{n-1}d\theta' = \left[\frac{(n-1){\theta'}^n}{n}\right]_0^\theta = \frac{(n-1)\theta^n}{n}.$$

Another way to derive this,

$$u_i(\theta) = \int_0^\theta u_i'(\theta')d\theta' = \int_0^\theta \Phi_{-i}(\theta')d\theta' = \int_0^\theta {\theta'}^{n-1}d\theta' = \frac{\theta^n}{n}$$

we also know that

$$u_i(\theta) = \theta\,\Pr(\text{win}) - b_i(\theta) = \theta^n - b_i(\theta)$$

So the two together imply that

$$b_i(\theta) = \frac{n-1}{n}\theta^n.$$

To check that this is consistent with the revenue equivalence theorem, note that

$$\mathrm{E}(Revenue) = n\int_0^1 \frac{n-1}{n}\theta^n d\theta = (n-1)\left[\frac{\theta^{n+1}}{n+1}\right]_0^1 = \frac{n-1}{n+1}.$$

- If we want to derive the bidding function the "stupid" way:

$$\max_x \theta\,\Pr\big(b(\theta_j) < x, \forall j\big) - x = \theta\,\Pr\big(\theta_j < b^{-1}(x), \forall j\big) - x = \theta\big(b^{-1}(x)\big)^{n-1} - x$$

FOC is

$$\theta(n-1)\big(b^{-1}(x)\big)^{n-2}\frac{db^{-2}(x)}{dx} = 1$$

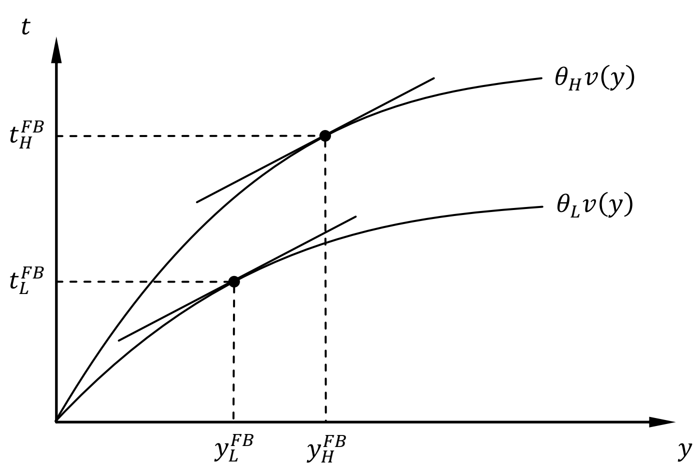
Impose symmetry, i.e. $x = b(\theta)$:

$$\theta(n-1)\theta^{n-2}\frac{1}{b'(\theta)} = 1 \;\Rightarrow\; b'(\theta) = (n-1)\theta^{n-1}$$

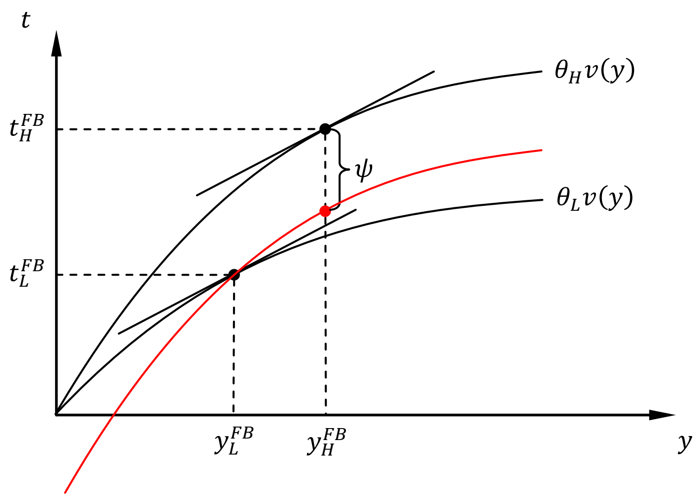
Optimal Mechanisms

- Given a set of implementable outcomes, what are the best mechanisms that implement these outcomes?
- Example. Monopolistic price discrimination.
  - 2 risk neutral parties: Principal P and Agent A
  - Outcome (y, t), where y is the quantity of good sold and t is the total price paid.
  - P's utility: $u_P(y, t) = t - cy$, where $c > 0$
  - A's utility: $u_A(y, t; \theta) = \theta v(y) - t$, where $v' > 0$, $v'' < 0$, $v(0) = 0$, $\lim_{y\to 0} v'(y) = \infty$, $\lim_{y\to\infty} v'(y) = 0$.
    - Note: u_A satisfies single-crossing:
    $$u_A(y, t; \theta_H) - u_A(y, t; \theta_L) = (\theta_H - \theta_L)v(y)$$
    which is strictly increasing in y.
  - 2 types: $\theta_L < \theta_H$. Let $\Pr(\theta = \theta_H) = p \in (0,1)$.

  - *Case* 1 (First best, full information). P's problem is
  $$\max_{y_H, t_H, y_L, t_L} p(t_H - cy_H) + (1 - p)(t_L - cy_L)$$
  subject to IR (individual rationality) constraints:
  $$\theta_L v(y_L) - t_L \geq 0 \quad \text{IR}_L$$
  $$\theta_H v(y_H) - t_H \geq 0 \quad \text{IR}_H$$
  Notice that the two constraints are independent of each other, and so are the two summands in the objective function. This allows us to maximize the following separately:
  $$\max_{t_H, y_H} t_H - cy_H, \qquad s.t.\ \theta_H v(y_H) - t_H \geq 0$$
  $$\max_{t_L, y_L} t_L - cy_L, \qquad s.t.\ \theta_L v(y_L) - t_L \geq 0.$$
  Note also that both constraints must be binding at the optimum, so that the FOC is
  $$\theta_i v'(y_i^{FB}) = c \;\Rightarrow\; t_i = \theta_i v(y_i^{FB}), \qquad i = H, L.$$



- *Case* 2 (Asymmetric information, P doesn't see θ)
  - Problem: θ_H prefers (y_L, t_L) to (y_H, t_H)
  - One solution: reduce t_H by ψ (aka the "information rent"). This implements the efficient outcome. But this doesn't maximize P's utility.



  - P can increase profit by selling less to the *low* type.
  - Second-best: the best that P can do subject to information limitation.

$$\max_{y_H, y_L, t_H, t_L} p(t_H - cy_H) + (1-p)(t_L - cy_L)$$

subject to

$$\theta_L v(y_L) - t_L \geq 0 \qquad \text{IR}_L$$
$$\theta_H v(y_H) - t_H \geq 0 \qquad \text{IR}_H$$
$$\theta_H v(y_H) - t_H \geq \theta_H v(y_L) - t_L \qquad \text{IC}_H$$
$$\theta_L v(y_L) - t_L \geq \theta_L v(y_H) - t_H \qquad \text{IC}_L$$

If all four constraints are binding, then there is no maximization problem and we'd just be solving the four constraints for the four variables. But based on intuition, we should expect that some of the four constraints are slack.

- $\text{IC}_H + \text{IR}_L \Rightarrow \text{IR}_H$ because

$$\theta_H v(y_H) - t_H \overset{\text{by IC}_H}{\geq} \theta_H v(y_L) - t_L \overset{\because \theta_H > \theta_L \ \&\ v(y_L) \geq 0}{\geq} \theta_L v(y_L) - t_L \overset{\text{by IR}_L}{\geq} 0$$

- (IC_H with equality) + ($y_H \geq y_L$) $\Rightarrow$ IC_L because

$$\left.\begin{array}{r} y_H \geq y_L \Rightarrow v(y_H) \geq v(y_L) \\ \theta_H v(y_H) - t_H = \theta_H v(y_L) - t_L \end{array}\right\} \Rightarrow \theta_L v(y_H) - t_H \leq \theta_L v(y_L) - t_L$$

Thus, we can replace IC_L with a "monotonicity" condition $y_H \geq y_L$.

So the maximization problem for the second best above is equivalent to

$$\max_{y_H, y_L} p(\theta_H v(y_H) - \theta_H v(y_L) + \theta_L v(y_L) - cy_H) + (1-p)(\theta_L v(y_L) - cy_L)$$

subject to $y_H \geq y_L$. The FOC w.r.t y_H is

$$\theta_H v'(y_H^{SB}) = c \ \Rightarrow\ y_H^{SB} = y_H^{FB}$$

FOC w.r.t y_L is

$$-p\theta_H v'(y_L^{SB}) + p\theta_L v'(y_L^{SB}) + (1-p)\theta_L v'(y_L^{SB}) = (1-p)c$$

$$\theta_L v'(y_L^{SB}) = \frac{c}{1 - \frac{p}{1-p} \cdot \frac{\theta_H - \theta_L}{\theta_L}} > c \quad \Rightarrow \quad y_L^{SB} < y_L^{FB}$$

Note that if $p\theta_H > \theta_L$ (i.e. the denominator above is negative), then $y_L^{SB} = 0$. Plug y_H^{SB}, y_L^{SB} back into the constraints to find the transfers:

$$t_L^{SB} = \theta_L v(y_L^{SB})$$

$$t_H^{SB} = \theta_L v(y_L^{SB}) + \theta_H [v(y_H^{SB}) - v(y_L^{SB})] = \theta_H v(y_H^{SB}) + \underbrace{(\theta_H - \theta_L) v(y_L^{SB})}_{\text{information rent}}$$

where $(\theta_H - \theta_L) v(y_L^{SB})$ is the information rent.

Monopolistic Screening with Continuous Types

- Environment
  - Type $\theta \in [\underline{\theta}, \bar{\theta}]$ with density $f(\theta)$ and cdf $F(\theta)$
  - Agent has quasi-linear utility $u(\theta, y, t) = \theta v(y) - t$, with $v' > 0$, $v'' < 0$, and $v(0) = 0$
  - Monopolist has constant MC, $c > 0$, and maximizes

$$\max_{t(\theta), y(\theta)} \int_{\underline{\theta}}^{\bar{\theta}} [t(\theta) - cy(\theta)] f(\theta) d\theta$$

subject to

$$\theta v(y(\theta)) - t(\theta) \geq 0, \qquad \forall \theta \qquad \text{(IR)}$$

$$\theta v(y(\theta)) - t(\theta) \geq \theta v(y(\theta')) - t(\theta'), \qquad \forall \theta, \theta' \qquad \text{(IC)}$$

We can reduce the number of IR constraints:

$$\theta v(y(\theta)) - t(\theta) \overset{\text{by IC}}{\geq} \theta v\left(y(\underline{\theta})\right) - t(\underline{\theta}) \overset{\substack{\theta > \underline{\theta} \\ v(\cdot) \geq 0}}{\geq} \underline{\theta} v\left(y(\underline{\theta})\right) - t(\underline{\theta}) \overset{\text{by IR}}{\geq} 0$$

Assuming that $y(\cdot), t(\cdot)$ are differentiable, we have the FOC (for truth telling):

$$\frac{\partial}{\partial \theta'} [\theta v(y(\theta')) - t(\theta')]\Big|_{\theta' = \theta} = 0$$

The SOC is

$$\theta v'(y(\theta)) y''(\theta) + \theta v''(y(\theta)) [y'(\theta)]^2 - t''(\theta) \leq 0$$

Want to replace the SOC with a monotonicity condition similar to the discrete case

$$\frac{\partial FOC}{\partial \theta} = v'(y(\theta)) y'(\theta) + \underbrace{\theta v''(y(\theta)) [y'(\theta)]^2 + \theta v'(y(\theta)) y''(\theta) - t''(\theta)}_{SOC} = 0$$

Since SOC is non-positive, it follows that

$$v'(y(\theta)) y'(\theta) \geq 0 \iff y'(\theta) \geq 0$$

Therefore, we can replace the IC constraints with

$$\theta v'(y(\theta)) y'(\theta) = t'(\theta), \qquad \forall \theta$$

$$y'(\theta) \geq 0, \qquad \forall \theta$$

Define

$$W(\theta) = \theta v(y(\theta)) - t(\theta) \Rightarrow W'(\theta) = v(y(\theta)) + \underbrace{\theta v'(y(\theta)) y'(\theta) - t'(\theta)}_{= 0 \text{ by IC}}$$

Then, by IR and $W(\underline{\theta}) = 0$,

$$t(\theta) = \theta v(y(\theta)) - W(\theta) = \theta v(y(\theta)) - \int_{\underline{\theta}}^{\theta} v(y(\theta')) d\theta'$$

The maximization problem becomes

$$\max_{y(\theta)} \int_{\underline{\theta}}^{\bar{\theta}} \left[\theta v(y(\theta)) - \int_{\underline{\theta}}^{\theta} v(y(\theta')) d\theta' - cy(\theta)\right] f(\theta) d\theta$$

$$= \max_{y(\theta)} \int_{\underline{\theta}}^{\bar{\theta}} [\theta v(y(\theta)) - cy(\theta)] f(\theta) d\theta - \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} v(y(\theta')) d\theta' f(\theta) d\theta$$

$$= \max_{y(\theta)} \int_{\underline{\theta}}^{\bar{\theta}} \left[\theta v\big(y(\theta)\big) - cy(\theta)\right] f(\theta) d\theta - \int_{\underline{\theta}}^{\bar{\theta}} \int_{\theta'}^{\bar{\theta}} f(\theta) d\theta v\big(y(\theta')\big) d\theta'$$

$$= \max_{y(\theta)} \int_{\underline{\theta}}^{\bar{\theta}} \left[\theta v\big(y(\theta)\big) - cy(\theta)\right] f(\theta) d\theta - \int_{\underline{\theta}}^{\bar{\theta}} [1 - F(\theta')] v\big(y(\theta')\big) d\theta'$$

subject to $y'(\theta) \geq 0$. Note that we could also use integration by parts to simplify the maximization problem:

$$\int_{\underline{\theta}}^{\bar{\theta}} \underbrace{\int_{\underline{\theta}}^{\theta} v\big(q(\theta')\big) d\theta'}_{u} \underbrace{f(\theta) d\theta}_{dv} = [uv]_{\underline{\theta}}^{\bar{\theta}} - \int_{\underline{\theta}}^{\bar{\theta}} v du$$

$$= \left[\int_{\underline{\theta}}^{\theta} v\big(q(\theta')\big) d\theta' \, F(\theta)\right]_{\underline{\theta}}^{\bar{\theta}} - \int_{\underline{\theta}}^{\bar{\theta}} F(\theta) v\big(q(\theta)\big) d\theta$$

$$= \int_{\underline{\theta}}^{\bar{\theta}} v\big(q(\theta)\big) d\theta - \int_{\underline{\theta}}^{\bar{\theta}} F(\theta) v\big(q(\theta)\big) d\theta$$

As in the discrete case, ignore the constraint and check consistency later. Maximizer pointwise

$$\max_{y(\theta)} \left[\theta v\big(y(\theta)\big) - cy(\theta)\right] f(\theta) - [1 - F(\theta)] v\big(y(\theta)\big), \qquad \forall \theta$$

The FOC is

$$\left[\theta v'\big(y(\theta)\big) - c\right] f(\theta) - [1 - F(\theta)] v'\big(y(\theta)\big) = 0$$

$$\left[\theta - \frac{1 - F(\theta)}{f(\theta)}\right] v'\big(y^{SB}(\theta)\big) = c$$

Compare to the first-best:

$$\theta v'\big(y^{FB}(\theta)\big) = c \;\; \Rightarrow \;\; y^{SB}(\theta) < y^{FB}(\theta), \qquad \forall \theta < \bar{\theta}$$

Monopolistic Screening with Continuous Types (cont'd)

- Recall from last time that the second best is given by

$$\left[\theta - \frac{1 - F(\theta)}{f(\theta)}\right] v'\left(q^{SB}(\theta)\right) = c$$

  Compare this with the first best

$$\theta v'\left(q^{FB}(\theta)\right) = c$$

  This means that

$$v'\left(q^{SB}(\theta)\right) > v'\left(q^{FB}(\theta)\right) \iff q^{SB}(\theta) < q^{FB}(\theta), \qquad \forall \theta < \bar{\theta}$$

  Under-consumption is most severe when
  - $\frac{1-F(\theta)}{f(\theta)}$ is big, i.e. when $f(\theta)$ and/or $F(\theta)$ is small.
  - A small $F(\theta)$ tells you how many people are below θ. If $F(\theta)$ is small, there is a lot to *gain* by under-providing to the low type, because the monopolist can charge more to the high types.
  - A small $f(\theta)$ tells you whether there is a lot of type θ people. If $f(\theta)$ is small, then the *loss* by under-providing to low types is small, because the loss of revenue from the low types is not that much.

- The sufficient condition (for replacing the constraints with FOC's and the monotonicity of q):

$$\frac{1 - F(\theta)}{f(\theta)} \text{ is decreasing with } \theta \iff \frac{f(\theta)}{1 - F(\theta)} \text{ is increasing with } \theta$$

  - In other words, we need a monotone hazard rate
  - This is coming from the constraint of monotonicity of q:

$$q'(\theta) \geq 0 \Rightarrow v'\left(q^{SB}(\theta)\right) \text{ must be decreasing with } \theta$$

$$\Rightarrow \theta - \frac{1 - F(\theta)}{f(\theta)} \text{ must be increasing with } \theta$$

Moral Hazard

- The standard environment:
  - Principal P owns technology $F(x, a)$ where
    - $x \in \{x_1, \dots, x_n | x_1 < \cdots < x_n\}$, is the outcome
    - $a \in A$, where A is the finite action set
    - $F(x, a)$ is the probability distribution over x given a, and
$$f_i(a) = \Pr(x = x_i | a) > a, \qquad \forall a \in A, \forall i \in \{1, \dots, n\}$$
  - P cannot perform task → must delegate to agent A
  - Action a is not observable (or verifiable), but the outcome x is observable and verifiable
  - P can offer contract where wage depends on x to induce a
  - Preferences:
$$u_P(x, a, w) = v(x) - w, \qquad \text{[doesn't depend on } a\text{]}$$
$$u_A(x, a, w) = u(w) - g(a), \qquad u' > 0, u'' \leq 0$$
- First best: action a is observable and verifiable
  - P can effectively pick the action for A by setting $w = -\infty$ if $a \neq a^*$
  - P solves, for any a,
$$\max_{w_1,\dots,w_n} \mathrm{E}[u_P] = \sum_i f_i(a)[v(x_i) - w_i]$$
subject to
$$\sum_i f_i(a)u(w_i) - g(a) \geq \bar{u} \quad \text{(IR)}$$
IR must be binding, for otherwise P can simply decrease the wages. The Lagrangian is
$$\mathcal{L} = \sum_i f_i(a)[v(x_i) - w_i] + \lambda\left[\sum_i f_i(a)u(w_i) - g(a) - \bar{u}\right]$$
The FOC is
$$-f_i(a) + \lambda f_i(a)u'(w_i) = 0 \;\; \Rightarrow \;\; u'(w_i) = \frac{1}{\lambda}, \qquad \forall i$$
This means that if $u(\cdot)$ is strictly concave ($u'' < 0$), then w_i is the same for all i.
$$w \equiv w_i \;\; \Rightarrow \;\; u(w) - g(a) = \bar{u} \;\; \Rightarrow \;\; w(a) = u^{-1}\big(\bar{u} + g(a)\big)$$
To find the optimal a, P solves
$$\max_{a \in A} \sum_i f_i(a)v(x_i) - w(a) \;\; \longrightarrow \;\; a^{FB}$$
  - Here, the agent bears no risk and the principal bears all the risk. This is because the principal is risk neutral, but the agent is risk averse.
  - Moral hazard:
    - If a is not observable/verifiable, the offering a fixed wage means that the agent will

pick $\text{argmin}_a\, g(a)$. This is inefficient when it is not equal to a^{FB}.

- If the principal offers variable wage, then the agent has to bear risk. This is inefficient when the agent is risk-averse and the principal is risk-neutral.

Moral Hazard (cont'd)

- Missed a lecture on March 29, 2012.
- Suppose A has only two actions: $a \in \{a_L, a_H\}$, and $g(a_H) > g(a_L)$
  - If P wants to induce a_H, then
$$v'(u_i) = \lambda + \mu_L \left(1 - \frac{f_i(a_L)}{f_i(a_H)}\right)$$
where $v = u^{-1}$.
  - Order the outcome x_i's by the index. When is w_i increasing in i?
$$w_i \uparrow \Leftrightarrow \; u_i \uparrow \Leftrightarrow \; RHS \uparrow \Leftrightarrow \; \frac{f_i(a_L)}{f_i(a_H)} \downarrow \Leftrightarrow \; \frac{f_i(a_H)}{f_i(a_L)} \uparrow$$
This is called the ***monotone likelihood ratio property*** (***MLRP***); that is, the likelihood ratio $f_i(a_H)/f_i(a_L)$ is increasing in i. An implication of this property is that $F(x, a_H) \succsim_{fsd} F(x, a_L)$.
- Characteristics of optimal contract
  - If $a \notin \text{argmin}_a \, g(a)$, then A is not fully insured, $\text{E}[w] > w^{FB}$
  - In the optimal scheme, wages don't directly depend on P's benefit from A's work. x_i only matter through the probabilities with which they occur.
    - The outcomes only matter through the probability, which P uses to statistically determine whether A has put in the effort or not.
  - IR is binding, so A gets no surplus.
- What if A is risk neutral? Let $u(w) = w$.
  - Consider this contract: $w_i = v(x_i) - K$, where K is a constant.
  - A solves the problem
$$\max_a \sum_i f_i(a)(v(x_i) - K) - g(a) = \max_a \sum_i f_i(a) v(x_i) - g(a)$$
But this is exactly the same maximization problem as in the first-best case.
  - P can then set $K = \sum_i f_i(a^{FB}) v(x_i) - w^{FB}$, which makes IR binding. So K is basically the price at which P sells the project to A.
  - A's expected utility is $w^{FB} - g(a^{FB}) = \bar{u}$, the reservation utility in the first-best case
    - In this case, A is the ***residual claimant***, because she is getting all the variation in profit, while P is getting a fixed amount. This is okay because A is risk neutral.
  - P's expected utility is $\sum_i f_i(a^{FB}) v(x_i) - w^{FB}$, also his first-best utility.
- What if the distribution is "funny"?
  - *Case* 1. a^{FB} is such that $\forall i$ where $f_i(a^{FB}) > 0$, $f_i(a) = 0 \;\; \forall a \neq a^{FB}$
    - In this case, P can pay w^{FB} if i is such that $f_i(a^{FB}) > 0$, and $-\infty$ otherwise, because

he can distinguish a^{FB} perfectly from any other actions a.

- This scheme is akin to dominant strategy implementation: choosing a^{FB} is always the dominant strategy.

> *Case* 2. $\exists i$ such that $f_i(a^{FB}) = 0$ and $f_i(a) > 0\ \forall a \neq a^{FB}$

- In this case, pay $-\infty$ if x_i, and w^{FB} otherwise. This works because if A chooses $a \neq a^{FB}$, there is a chance that x_i happens and she'd get $-\infty$.
- This scheme is like Bayesian implementation: it is only in expectation that choosing a^{FB} is optimal, but not in every case.

> *Case* 3. Suppose $x \sim \mathcal{N}(\mu(a), \sigma^2)$, $\mu', g' > 0$, $g'' > 0$. Want to induce a^{FB}. Consider the wage scheme:

$$w = \begin{cases} w^{FB} + \epsilon & \text{if } x \geq \underline{x} \\ -K & \text{if } x < \underline{x} \end{cases}$$

We want to show that ϵ can be made arbitrarily close to zero.

- Here, ϵ is in place to make sure that the IR constraint is satisfied.
- Fix a large K, look at by how much the IR would be violated if $\epsilon = 0$
- The IC constraint is

$$\underbrace{\int_{-\infty}^{\underline{x}} u(-K) f_a(x, a^{FB}) dx + \int_{\underline{x}}^{\infty} u(w^{FB}) f_a(x, a^{FB}) dx}_{=\frac{d}{da}\int_{-\infty}^{\infty} u(w) f(x, a^{FB}) dx} = g'(a^{FB})$$

- Violate IR by

$$\int_{-\infty}^{\underline{x}} [u(w^{FB}) - u(-K)] f(x, a^{FB}) dx$$

- Recall that we have a normal distribution:

$$f(x,a) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu(a)}{\sigma}\right)^2} \Rightarrow f_a(x,a) = f(x,a)\left(\frac{x-\mu(a)}{\sigma^2}\right)\mu'(a)$$

$$\Rightarrow \frac{f_a(x,a)}{f(x,a)} = \frac{1}{\sigma^2}\big(x - \mu(a)\big)\mu'(a) \xrightarrow{x \to -\infty} -\infty$$

This means that

$$\forall M > 0, \exists x, \forall x < \underline{x} : \frac{f_a(x,a)}{f(x,a)} < -M$$

So IR is violated by less than

$$-\frac{1}{M}\int_{-\infty}^{\underline{x}} [u(w^{FB}) - u(-K)] f_a(x, a^{FB}) dx \overset{\text{by IC}}{=}$$

$$-\frac{1}{M}\left[\int_{-\infty}^{\infty} u(w^{FB}) f_a(x, a^{FB}) dx - g'(a^{FB})\right]$$

$$= -\frac{1}{M}\underbrace{\left[\int_{-\infty}^{\infty} u(w^{FB}) \frac{x - \mu(a^{FB})}{\sigma^2} \mu'(a^{FB}) f(x,a) dx - g'(a^{FB})\right]}_{<\infty} \xrightarrow{M \to \infty} 0$$

- Optimal contracts all have this similar flavor: penalty happens very infrequently, but is very harsh.
- The property of the normal distribution is that if you're in the tail of the distribution, if effort level goes down, it makes a low outcome a lot more likely to happen.

Rubinstein Bargaining Model

- Environment
  - Two players: 1 and 2
  - Bargain over a "pie" of size 1. The outcome is $(u_1, u_2) = (x, 1-x)$ where $x \in [0,1]$
  - Discount factor $\delta < 1$
  - The game goes as follows:
    - *Period* (phase) 1. Player 1 proposes a division
      - If player 2 accepts, game ends, and proposal implemented.
      - If player 2 rejects, game goes to phase 2.
    - *Period* (phase) 2. Player 2 proposes a division
      - If player 1 accepts, game ends, and proposal implemented.
      - If player 1 rejects, game goes to phase 1.
    - If no agreements in the last period (if there is one), then both get $(0,0)$.
- Suppose game has T periods, and T is even.
  - In period T, in SPE, player 2 proposes $(0,1)$, and player 1 accepts
  - In period $T-1$, in SPE, player 1 proposes $(1-\delta, \delta)$, and player 2 accepts
  - In period $T-2$, in SPE, player 2 proposes $\big(\delta(1-\delta), 1-\delta(1-\delta)\big)$, and player 1 accepts
  - In period $T-3$, in SPE, player 1 proposes $\Big(1-\delta(1-\delta), \delta\big(1-\delta(1-\delta)\big)\Big)$, and player 2 accepts
  - In period 1, in SPE, player 1 proposes

$$\left(\underbrace{1-\delta+\delta^2-\delta^3+\delta^4-\delta^5+\cdots+\delta^{T-2}-\delta^{T-1}}_{u_1}, \underbrace{\delta-\delta^2+\delta^3-\delta^4+\cdots-\delta^{T-2}+\delta^{T-1}}_{u_2}\right)$$
$$= \left(\frac{1-\delta^T}{1+\delta}, 1-\frac{1-\delta^T}{1+\delta}\right)$$

- Suppose game is infinite ($T \to \infty$), the same logic applies

$$T \to \infty \;\; \Rightarrow \;\; (u_1, u_2) = \left(\frac{1}{1+\delta}, \frac{\delta}{1+\delta}\right)$$

  - Note that in general, we cannot go from finite to infinite in this way.
  - In this particular case, though, this is true. Here's why:
  - Let $\bar{v}_1$ be player 1's (the proposer) payoff in the best SPE
    - (more formally, $\bar{v}_1$ is the supremum of player 1's payoff in SPE's)

- Let $\underline{v}_1$ be player 1's (the proposer) payoff in the worst SPE (i.e. the infimum)
- By the recursive nature of the game, we deduce
$$\bar{v}_1 \leq 1 - \delta \underline{v}_1$$
where $\delta \underline{v}_1$ is player 2's lowest reservation utility, because he gets at least $\underline{v}_1$ in the next period. For the same reason, we have
$$\underline{v}_1 \geq 1 - \delta \bar{v}_1$$
Plug the second inequality into the first,
$$\bar{v}_1 \leq 1 - \delta(1 - \delta \bar{v}_1) = 1 - \delta + \delta^2 \bar{v}_1$$
$$(1 - \delta^2)\bar{v}_1 \leq 1 - \delta$$
$$\bar{v}_1 \leq \frac{1}{1 + \delta}$$
Plug the first inequality into the second,
$$\underline{v}_1 \geq 1 - \delta(1 - \delta \bar{v}_1) = 1 - \delta + \delta^2 \bar{v}_1$$
$$\underline{v}_1 \geq \frac{1}{1 + \delta}$$
Therefore, the only possible payoff for player 1 is $\frac{1}{1+\delta}$.
- So far we've shown that in SPE, $\bar{v}_1 = \underline{v}_1$, but still need to establish (unique) existence. This is simple:
  - In each period, proposer demands $\frac{1}{1+\delta}$ and offers $\frac{\delta}{1+\delta}$, receivers accepts.

Rubinstein Bargaining Model (cont'd)

- Allowing for different discount factors
  - Let $\bar{v}_i$ be the supremum of all SPE's of player i's payoff when he/she proposes in the first round; and similarly, let $\underline{v}_i$ be the infimum of all SPE's of player i's payoff when he/she proposes in the first round.

$$\left.\begin{aligned} \bar{v}_i &\leq 1-\delta_j \underline{v}_j \\ \underline{v}_j &\geq 1-\delta_i \bar{v}_i \end{aligned}\right\} \Rightarrow \begin{cases} \bar{v}_i \leq 1-\delta_j(1-\delta_i \bar{v}_i) \;\Rightarrow\; \bar{v}_i \leq \dfrac{1-\delta_j}{1-\delta_i\delta_j} \\ \underline{v}_j \geq 1-\delta_i(1-\delta_j \underline{v}_j) \;\Rightarrow\; \underline{v}_j \geq \dfrac{1-\delta_i}{1-\delta_i\delta_j} \end{cases}$$

  By symmetry, we can verify

$$\bar{v}_j \leq \frac{1-\delta_i}{1-\delta_i\delta_j} \quad \text{and} \quad \underline{v}_i \geq \frac{1-\delta_j}{1-\delta_i\delta_j}$$

  Using the argument from last time, we can show that

$$v_i = \frac{1-\delta_j}{1-\delta_i\delta_j}, \qquad v_j = \frac{1-\delta_i}{1-\delta_i\delta_j}$$

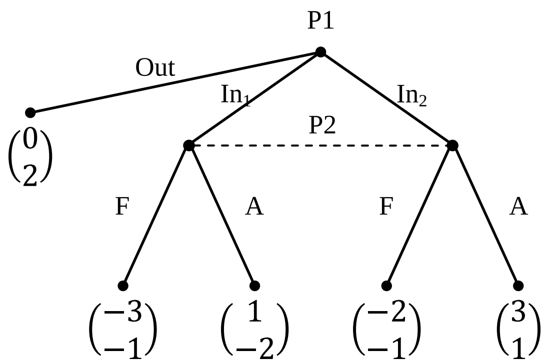
  If i proposes first, then he'll propose v_i for himself, and $1-v_i$ for j, which is

$$1-v_i = 1-\frac{1-\delta_j}{1-\delta_i\delta_j} = \delta_j \frac{1-\delta_i}{1-\delta_i\delta_j} = \delta_j v_j$$

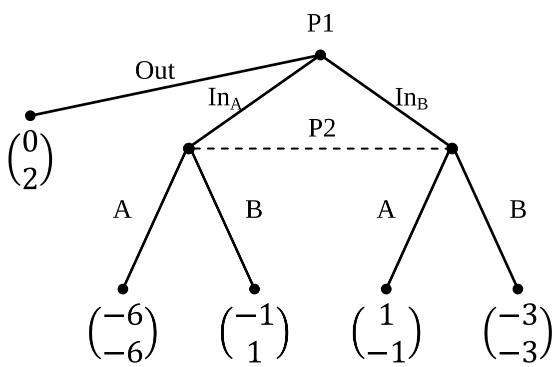
  - This shows that the more patient a player is, the higher his/her payoff will be.

Forward Induction

- Based on what happens earlier, a player might be able to infer some information that's not available previously.
- Consider a market entry game



  - Consider (Out, F).
    - This is a SPE.
    - But this is not reasonable, because P2 knows that P1 will never choose In_1 if it chooses to go in at all: because In_2 strictly dominates In_1.
    - If P1 chooses In_2, P2's best response is A.
- Consider another market entry game



  - Consider (Out, B)
    - This is an SPE.
    - If P2 chooses B, it means that she places a higher probability of P1 playing In_A.
    - But Out dominates In_A. Thus, it is unreasonable for P2 to expect that P1 to play In_A.
    - P2's best response to In_B is A.
- Formally, the criterion for forward induction is that one should place a belief of zero on dominated actions.

- Note that this is stronger than both weak PBE and sequential equilibrium. In the two examples given, the unreasonable equilibria are both weak PBE and SE.