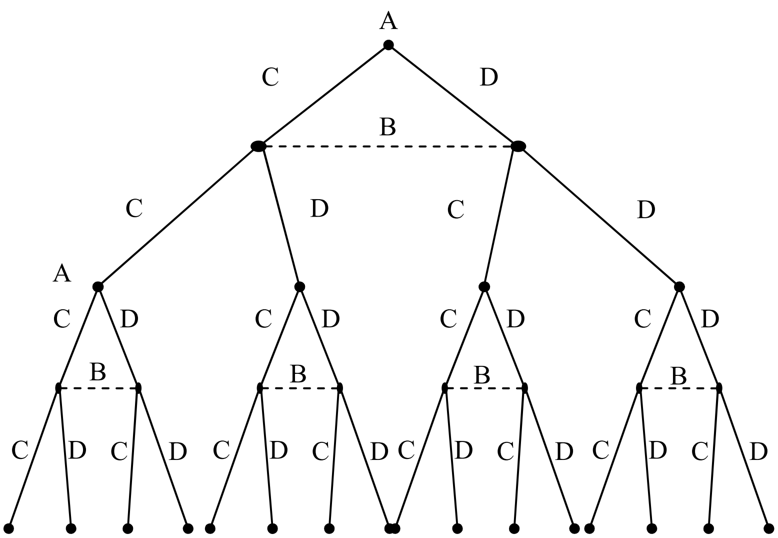
Review of Repeated Games

- Repeated Game
  - Perfect Information
  - Objective: Can threats or promises about future behavior influence current behavior in repeated relationships?

- Example. Prisoners' Dilemma

	C	D
C	4, 4	0, 5
D	5, 0	1, 1

  - Repeating this game:



    - NE is unique (so is SPNE), because defect is strictly dominant in each subgame.

- Notations
  - Let $G = \{A_1, \dots, A_n, u_1, \dots, u_n\}$ be a static game of complete information in which players $1, \dots, n$ choose actions $a_i \in A_i$ simultaneously, and payoffs are $u_1(a_1, \dots, a_n), \dots, u_n(a_1, \dots, a_n)$. Call G a static game.
  - Let G be repeated $t = 0, \dots, T$ times ($T = \infty$ is possible)
  - Pure strategies only.
    - Let a_i^t be player i's action at t, $a_i \in A_i$
    - Let $a^t = (a_1^t, \dots, a_n^t)$ be the action profile at t
    - Let $h^t = \{a^0, a^1, \dots, a^{t-1}\}$ be the set of possible histories
    - Let $s_i^t : H^t \to A_i$ be the strategy function for player i that maps the set of possible histories into the set of actions

- Finitely Repeated Games (T is finite)
  - **Proposition.** If G has a unique NE, then the finitely repeated game G^T has a unique SPE in which the NE of G is played at every stage $t = 1, \dots, T$.

Repeated Games (cont'd)

- Prisoners' Dilemma repeated twice (with perfect recall)

	C	D
C	4, 4	0, 5
D	5, 0	1, 1

- **Proposition.** If the stage game G has a unique Nash equilibrium (NE), then for any finite T, the repeated game G^T has a unique subgame perfect equilibrium, in which the unique NE of G is played in all $t = 0, \ldots, T$.
  - *Proof.* Use backward induction. In T, there is no future, so the unique NE will be played. That means, in $T - 1$, there is no future to be conditioned on, so the unique NE will again be played.

  - Note that the prisoners' dilemma game has the property that its NE is the "minmax" payoff.
    - *Definition.* A player's ***minmax payoff*** is given by
$$v_i = \min_{a_{-i}} \max_{a_i} u_i(a_i, a_{-i})$$
      - Note that this may be due to the consideration that the other players are trying to "punish" player i.
      - Each game has a minmax payoff.
    - Example of minmax payoff.

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

Player 1's feasible payoffs are -2, 0, 1. Note that Player 1 can play D, such that his minmax payoff is at least 0. But can Player 2 hold Player 1 to playing D? Apparently no pure strategy of Player 2 can do this. If we allow for mixed strategy, suppose Player 2 plays L with p and R with $1 - p$. Then,
$$v_D = 0$$
$$v_M = p - 2(1 - p)$$
$$v_U = -2p + (1 - p)$$
$$v_M = v_U \;\Rightarrow\; p - 2(1 - p) = -2p + (1 - p) \;\Rightarrow\; p = \frac{1}{2}$$

- **Proposition 2.** If the payoff profile in every NE of G is the minmax profile, then for any T, the outcome $(a^0, \ldots, a^T)$ of any NE of G^T is such that a^t is a NE of G for all $t = 1, \ldots, T$.
  - In other words, if there is a unique NE with minmax payoffs, then G^T has unique NE with minmax payoffs.
    - *Proof.* By contradiction. Consider the prisoners' dilemma.
The minmax strategy is
$$(D, D, D, D, D, \ldots, D)$$
Suppose there is another NE strategy
$$(D, D, C, C, C, \ldots, D)$$

- Observation. In any NE of G^T the average payoffs are at least equal to the minmax payoffs of the stage game G.

- Suppose we have multiple NE—what happens then?

	C	D	R
C	4, 4	0, 5	0, 0
D	5, 0	1, 1	0, 0
R	0, 0	0, 0	3, 3

Possible SPNE: (D, D regardless), (R, R regardless) (D, R regardless) (R, D regardless)
These all consist of NE in the stage game. But the following is also SPNE:

$$\left(C, \quad \begin{cases} R & \text{if } C \\ D & \text{otherwise} \end{cases}\right)$$

- Recall (notations)

$a^t = (a_1^t, \dots, a_N^t)$, action profile in t, $a_i \in A_i$
$h^t = (a^0, a^1, \dots, a^{t-1})$, history at t
$s_i^t : H^t \to A_i$
$s_i = \{s_i^t\}_{t=0}^T$

Assume: plays max either discounted sum of payoffs:

$$\sum_{t=0}^{T} \delta^t u_i(a^t)$$

or average discounted present value

$$\frac{1-\delta}{1-\delta^{T+1}} \sum_{t=0}^{T} \delta^t u_i(a^t)$$

Repeated Games (cont'd)

- Recall that when the prisoners' dilemma game is repeated twice, we could get people to play the Pareto superior strategy at the first stage.
  - Benoit/Krishna (1985, *Econometrica*)

Infinitely Repeated Games

- We don't have the last period—so we cannot use backward induction.

- ***One Stage Deviation Principle***.
  - *Definition.* A ***one stage deviation*** from a strategy s_i is a strategy $\hat{s}_i$ such that

$$\hat{s}_i(h_t) \neq s_i(h_t), \qquad \text{for some unique } t \text{ and } h_t$$
$$\hat{s}_i(\tilde{h}_t) = s_i(\tilde{h}_t), \qquad \forall \tilde{h}_t \neq h_t$$

  - **Proposition 3. *One Stage Deviation Principle*** (for multi-stage games with observed actions) A strategy profile $s = (s_1, \ldots, s_N)$ is subgame perfect if and only if there is no profitable one-stage deviation after any history $h_t \in H_t$ for any player $i \in \{1, \ldots, N\}$.
    - In other words, no one can gain from deviating once and conforming to s thereafter.

    Proof. (⇐) "obvious": if it is profitable to deviate once, then it's not subgame perfect.

    (⇒) it's easier to prove the contrapositive: if a strategy profile is not subgame perfect, then it is profitable to deviate at least once.

    If s_i is not SPE, then there exists a deviation $\tilde{s}_i \neq s_i$ which is profitable for i. If the deviation happens in T periods, then it is also profitable to just deviate in one period.

    Consider a one stage deviation that occurs in the "last" period T,

$$s_i^T(h_t) = \begin{cases} \tilde{s}_i(h_t) & \text{if } h_t = \tilde{h}_T \\ s_i(h_t) & \text{if } h_t \neq \tilde{h}_T \end{cases}$$

    Two possibilities:
    - If $s_i^T(h_t)$ is profitable, then the one stage deviation is profitable → done!
    - If $s_i^T(h_t)$ is not profitable, then look at s_i^{T-1}, and so on. By iterative argument, there must be <u>one</u> one-stage deviation.
    - To prove the proposition for $T = \infty$, need "payoffs continuous at infinity" (i.e. payoffs are bounded).

Folk Theorem

- Assume
  - 2 players
  - Stage game G has a unique NE a^*

- *Definition.* A strategy profile s^* in a repeated game G^∞ involves ***Nash Reversion*** if and only if s_i calls for playing some outcome path $\{a^t\}_{t=0}^{\infty}$ until one player deviates, when a^* is played thereafter.

- **Lemma.** A Nash reversion strategy profile s^* form a SPE if and only if

$$\forall t, \forall \tilde{a}_i \in A_i, \forall i: \sum_{\tau=t}^{\infty} \delta^{\tau-t} u_i(a^\tau) \geq u_i(\tilde{a}_i, a_{-i}) + \underbrace{\sum_{\tau=t+1}^{\infty} \delta^{\tau-t-1} u_i(a^*)}_{=\frac{\delta}{1-\delta}u_i(a^*)}.$$

  - *Proof.* (⇐) Suppose the above inequality does not hold. Then, the inequality sign reverts, and thus there exists a profitable one-shot deviation. Thus, s^* is not SPE.
(⇒) Suppose s^* is not SPE. Then there exists a profitable one-shot deviation, and thus the above inequality does not hold (note that the above inequality says that there does not exist a profitable one-shot deviation). ■

  - Note 1. The above inequality greatly simplify if the outcome path is stationary, i.e. $a^t = a \ \forall t$. That is,

$$\max_{\tilde{a}_i} u_i(\tilde{a}_i, a_{-i}) - u_i(a) \leq \frac{\delta}{1-\delta}[u_i(a) - u_i(a^*)], \qquad i = 1,2$$

  - Note 2. The above inequality is easier to satisfy the higher the δ. That is, if the inequality is satisfied for δ, then it also holds for $\delta' > \delta$.
    - If $\delta \to 1$, everything that is better than the NE of G is supported.

- **Proposition 4 (Folk Theorem I due to Friedman (1971)).**
Let $a \in A$ be the stage game action profile such that

$$u_i(a) > u_i(a^*), \qquad i = 1,2.$$

Then, there exists a $\underline{\delta}$ such that whenever $\delta \geq \underline{\delta}$, playing a in every period t constitutes a SPE outcome path with Nash reversion strategies.
  - Issue 1. There is a possibility of re-negotiation.
  - Issue 2. The NE could involve very high payoffs, so that the "punishment" is not harsh enough. If the players' minmax payoff is lower than the NE payoff, then the "punishment" can be made harsher by playing the minmax strategies (but this leads to the problem whether the player has an incentive to play the minmax strategy). So do not use Nash reversion *unless* it is also the minmax.
    - This is to say that, if there is a NE whose payoff profile coincides with the minmax payoff profile, then any outcome that can be supported by any SPE can also be supported by Nash reversion strategies. That is, it is *sufficient* in such cases to consider Nash reversion equilibria in characterizing the set of payoffs that can be supported as the average payoff of some SPE. However, if no NE whose payoff profile coincides with the minmax payoff profile, then the set of Nash reversion

equilibria may exclude some equilibrium outcomes; namely, there may be other outcomes that can be supported by some SPE but not Nash reversion equilibrium. (cf. Mailath & Samuelson (2006, 77))

- Recall:

$$\underline{v}_i = \min_{a_{-i}} \left\{ \max_{a_i} u_i(a_i, a_{-i}) \right\}$$

- *Definition.* The set of feasible, individually rational payoffs (convexified) is

$$V^{IR} = convex\ hull\{v \geq \underline{v}_i | \forall i, \exists a : u_i(a) = v\}.$$

- **Proposition 5 (Folk Theorem II, due to Fudenberg & Maskin (1986)).**
  Suppose that $\dim(V^{IR}) = 2$. For any $v \in V^{IR}$ [note that v has to be STRICTLY individually rational], there exists a $\underline{\delta} < 1$ such that for all $\delta \geq \underline{\delta}$, there is a SPE of G^{∞} with average payoff equal v.
  - Problem. Standard equilibrium concepts do not pin down the path of play of patient players.
  - *Proof.* The basic idea is to have players minmax a deviating player and reward them in later periods for punishing the deviator.
    Assume that m^i is the minmax action profile is a pure strategy. Choose

    $$v' \in \text{Int}(V^{IR}) : v_i' < v_i, \qquad \forall i$$

    and let w^i denote v_i' with ϵ added to opponents payoff, i.e.

    $$w^i = (v_i', v_{-i}' + \epsilon)$$

    Strategies:
    - *Phase 1.* Play a (i.e. $u(a) = v$), as long as there are no deviations. If i deviates, switch to Phase 2_i.
    - *Phase 2_i.* Play m^i for T periods. If player j deviates (from minmaxing i), switch to Phase 2_j. If there are no deviations, then play moves to Phase 3. After T periods,
    - *Phase 3.* Play action profile leading to w^i forever. If j deviates, again, go to Phase 2_j.

Folk Theorem (cont'd)

- **Proposition 5 (Folk Theorem II, due to Fudenberg & Maskin (1986)).**
  - *Proof* (continue from last time). Given the strategies defined, choose
$$\underline{v}_i < v_i' < v_i \;\;\Rightarrow\;\; w^i = (v_i', v_{-i}' + \epsilon).$$
Choose T such that (for $\delta \to 1$)
$$\underbrace{\overbrace{\max_a u_i(a)}^{\bar{M}}}_{\substack{\text{upper bound}\\ \text{of payoff}}} + T\underline{v}_i < \underbrace{\overbrace{\min_a u_i(a)}^{\underline{M}}}_{\substack{\text{lower bound}\\ \text{of payoff}}} + Tv_i'$$
Need to check that players don't have incentive to deviate in each Phase.
In Phase 1,
$$v_i > (1-\delta)\bar{M} + \delta(1-\delta^T)\underline{v}_i + \delta^{T+1}v_i'$$
hold for $\delta \to 1$. So there is no profitable deviation in this period.

In Phase 3_j, $j \neq i$,
$$v_i' + \epsilon > (1-\delta)\bar{M} + \delta(1-\delta^T)\underline{v}_i + \delta^{T+1}v_i'$$
In Phase 3_i, $j = i$,
$$v_i' > (1-\delta)\bar{M} + \delta(1-\delta^T)\underline{v}_i + \delta^{T+1}v_i'$$
These two hold for $\delta \to 1$.

In Phase 2_j, player $i \neq j$ gets
$$\underbrace{(1-\delta^T)u_i(m^i) + \delta^T(v_i + \epsilon)}_{\text{conforming to minmaxing } j} > \underbrace{(1-\delta)\bar{M} + \delta(1-\delta^T)\underline{v}_i + \delta^{T+1}v_i'}_{\substack{\text{deviating from minmaxing } j\\ \text{and will be minmaxed by } j \text{ in the next period}}}$$

Dynamic (Stochastic) Games

- In repeated games: the physical environment is the same each period t (i.e. the stage game is stationary over time).
  - My action in this period affects the action tomorrow. However, today's action does not change the form of the game tomorrow (so dynamic game is to make an extension in this direction).

- Dynamic games:
  - Environment in each period by "state"
  - Current payoffs depend only on actions today and on the state
  - State follows a Markov process

- *Definition.* A ***dynamic*** (***stochastic***) ***game*** is
  - A set of players: $i = 1, \dots, N$
  - A set of actions: $a_i \in A_i(K^t),\ \forall i$, with $a^t = (a_1^t, \dots, a_N^t)$ being the action profile in t
  - Instantaneous utility function:

$$u_i(a^t, K^t) = E\left\{\sum_{s=0}^{\infty} \delta^s u_i(a^{t+s}, K^{t+s})\right\}, \qquad K^t \text{ is the state in } t$$

  - Transition function (law of motion):

$$q(K^{t+1} | a^t, K^t)$$

Dynamic Games (cont'd)

- *Definition.* Dynamic Games:
  - Set of players: $i = 1, \dots, N$
  - Set of actions: $a_i \in A_i(k),\ \forall i$
  - Instantaneous utility: $u_i(a^t, k^t)$

$$U_i^t = E\left[\sum_{\tau=0}^{\infty} \delta^\tau u_i(a^{t+\tau}, k^{t+\tau})\right]$$

  - Transition function: $q(k^{t+1}|a^t, k^t)$
  - Focus on games with observable actions (or perfect monitoring), i.e. players observes all past actions and the realization of k^t)
  - Public history at t: $h^t = (a^0, k^0, a^1, k^1, \dots, a^{t-1}, k^{t-1})$
  - Pure strategy at t: $s_i^t : H^t \times K \to A_i$

- *Definition.* A strategy profile $\hat{s}$ is a (stationary) ***Markov strategy profile*** if, for any two histories, h^t and $\tilde{h}^t$, of the same length (or of different length—this has to do with stationarity) and resulting in the same state k^t, we have $\hat{s}(h^t) = \hat{s}(\tilde{h}^t)$.
  - This embodies the idea that "bygones are bygones"
  - The Markov strategy rules out strategies like "if you play x last period, I'll play y today", because past histories does not matter.

- *Definition.* A strategy profile $\hat{s}$ is a ***Markov perfect equilibrium*** (***MPE***) if it is a SPE and $\hat{s}$ is a Markov strategy.
  - This is used to refine the SPE concept. The set of Markov perfect equilibria is a subset of the set of SPE's.

  - Notes.
    - The MPE is a refinement of SPE. (this is appealing because it typically reduces the number of equilibria).
    - Another appealing feature of MPE is that it makes simulation and estimation very easy.
    - Can also think of this as the simplest form of behavior that is consistent with rationality.
    - In finitely repeated games, if G has a unique NE, then there is a unique MPE in which this NE is played every period.

- Example (from Mailath & Samuelson). Common Pool Problem / Resource Extraction
  - Two players, $i = 1,2$
  - Extract resource from a common pool
  - Utility functions are

$$U_i^t = \sum_{\tau=0}^{\infty} \delta^\tau \ln(c_i^{t+\tau}),$$

  where c_i^t = consumption $\cong$ resource extracted at time t by player i
  - The stock of resources k^t evolves according to

$$k^{t+1} = 2(k^t - c_1^t - c_2^t), \qquad k^0 = initial\ state > 0, \qquad \forall t : k^t \geq 0$$

- Stage game: in each t
  - Players simultaneous announce c_i^t
  - Consume

$$c_i^t = \begin{cases} c_i^t & \text{if } c_1^t + c_2^t \leq k^t \\ \frac{1}{2}k^t & \text{if otherwise} \end{cases}$$

  - Use dynamic programming to solve the game. Look for stationary (independent of t) and symmetric (independent of i) MPE's: $c_i^t(k^t)$
    - Player i's value function:

$$V(k) = \max_{\tilde{c}} \left\{ \ln(\tilde{c}) + \delta V\left(2\big(k - \tilde{c} - c(k)\big)\right)\right\}$$

FOC's:

$$\frac{1}{c(k)} = 2\delta V'\left(2\big(k - \tilde{c} - c(k)\big)\right)$$

Use guess and verify: $c(k) = \alpha k$

$$c^0(k^0) = \alpha k^0$$
$$c^1(k^1) = \alpha k^1 = 2\alpha(k^0 - 2\alpha k^0) = 2\alpha k^0(1 - 2\alpha)$$
$$\vdots$$
$$c^t(k^t) = \alpha 2^t (1 - 2\alpha)^t k^0$$

Then the value function (with equilibrium strategies) is

$$V(k) = \sum_{\tau=0}^{\infty} \delta^\tau \ln(\alpha 2^\tau (1 - 2\alpha)^\tau k)$$

$$\Rightarrow \quad V'(k) = \sum_{\tau=0}^{\infty} \frac{\delta^\tau}{k} = \frac{1}{1-\delta} \cdot \frac{1}{k}$$

Plug into FOC:

$$\frac{1}{\alpha k} = 2\delta \frac{1}{1-\delta} \cdot \frac{1}{2(k - 2\alpha k)} \quad \Rightarrow \quad 2(1 - 2\alpha)(1 - \delta) = 2\delta\alpha$$

$$\Rightarrow \quad \alpha = \frac{1-\delta}{2-\delta} \leq \frac{1}{2}$$

Interpretation: in each t, players extract a proportion of 2α stock, leaving

$$1 - 2\alpha = 1 - 2\left(\frac{1-\delta}{2-\delta}\right) = \frac{\delta}{2-\delta}$$

of the stock for the next period, which then doubles. So

$$growth\ rate = \frac{2\delta}{2-\delta} \gtreqless 1, \qquad as\ \delta \gtreqless \frac{2}{3}$$

    - Compare this to the efficient consumption/resource extraction path: need to solve the planner's problem:

$$u_1 + u_2 = \sum_{t=0}^{\infty} \delta^t \{\ln(c_1^t) + \ln(c_2^t)\}$$

Note that $c_1^t = c_2^t$ by concavity of the log-utility. Value function for the planner is

$$V(k) = \max_{\tilde{c}}\left\{2\ln\tilde{c} + \delta V\big(2(k - 2\tilde{c})\big)\right\}$$

FOC:

$$\frac{1}{c(k)} = 2\delta V'\left(2\big(k - 2c(k)\big)\right)$$

Social Choice Theory

- Motivations:
  - Normative: efficient allocations from general equilibrium theory do not imply that they are *just*. We need an additional criterion (other than efficiency) to evaluate social outcomes. Social choice theory attempts to study the aggregation of individual preferences.
  - Positive: understanding collective choice process (e.g. ordering toppings on pizzas, and policy decisions, elections, etc.).

- Notations.
  - Set of alternatives: X (e.g. candidates of election, pizza toppings, etc.)
  - Set of individuals: $\mathcal{I} = \{1, \dots, I\}$
  - Individual preferences: $\succcurlyeq_i$ on X can be represented by $u_i(X)$
  - Set of all weak preference orderings on X: $\mathcal{R}$
    - Goal: social preference ordering $\succcurlyeq_s$

- *Definition.* A ***social welfare functional*** is a function

$$F : A \to \mathcal{R}$$

where $A \subseteq \mathcal{R}^I$ that assigns a social preference relation $\succcurlyeq_s \in \mathcal{R}$ to any profile of individual preference orderings $(\succcurlyeq_1, \dots, \succcurlyeq_I)$ on the domain $A \subseteq \mathcal{R}^I$.
  - Note.
    - We're taking <u>all</u> individuals' preferences as inputs.
    - We don't worry about how we know $\succcurlyeq_i$ (these are the true preferences).
    - We only consider ordinal preferences (i.e. intensities don't matter, neither does expertise)

- The case of two alternatives.
  - Outcomes: $X = \{x, y\}$, where $x =$ status quo, $y =$ reform.
  - Individuals: $i = \{1, \dots, I\}$

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

  - Preference profile is a list $(\alpha_1, \dots, \alpha_I)$

- *Definition.* A ***simple majority welfare functional*** is the function

$$F : \{+1, 0, -1\}^I \to \{+1, 0, -1\}$$

where $F = sign[\sum_i(\alpha_i)]$.
  - Don't confuse with *absolute majority*.
    - Simple majority can happen where everyone except one is indifferent. But absolute majority requires that over half of the population prefer one alternative over the other.

- Question: What are the characteristics of this simple majority rule?
  1. ***Universal Domain*** (***UD***): F assigns an unambiguous social ranking to all conceivable individual preference profiles.
  2. ***Symmetry*** or ***Anonymity*** (***S***): F treats individuals equally, i.e. permuting agents'

preferences doesn't alter the social ranking. (note that this rules out dictatorship)

3. ***Neutrality*** (***N***): F treats alternatives equally, i.e. reversing everybody's preferences reverses the social ranking:
$$F(-\alpha_1, \dots, -\alpha_I) = -F(\alpha_1, \dots, \alpha_I)$$
4. ***Positive Responsiveness*** (***PR***): if
$$(\alpha_1, \dots, \alpha_I) \geq (\alpha_1', \dots, \alpha_I')$$
with strict inequality for some i,
$$F(\alpha_1', \dots, \alpha_I') \geq 0 \;\; \Rightarrow \;\; F(\alpha_1, \dots, \alpha_I) = 1.$$
For example, the absolute majority rule does not satisfy PR, neither does the constant rule.

- **Theorem I (May (1952)).** A social welfare functional satisfies UD, S, N, PR if and only if it is the simple majority social welfare functional.
  - Note: MWG incorporates the UD into the definition of social welfare functional.

Proof of May's Theorem

- *Proof of May's Theorem.*
  - We already know that the simple majority social welfare functional satisfies UD, S, N, PR. So the ($\Rightarrow$) direction is done.
  - ($\Leftarrow$) Notice that
    - S $\Rightarrow \succcurlyeq_s$ only depends on $n^+ = \#\{i : \alpha_i = +1\}$ and $n^- = \#\{i : \alpha = -1\}$.
$$F(\alpha_1, \dots, \alpha_I) = G[n^+(\alpha_1, \dots, \alpha_I), n^-(\alpha_1, \dots, \alpha_I)]$$
    - Claim: $n^+(\alpha) = n^-(\alpha) \;\Rightarrow\; F(\alpha) = 0$
$$\begin{aligned} F(\alpha) &= G\big(n^+(\alpha_1, \dots, \alpha_I), n^-(\alpha_1, \dots, \alpha_I)\big) \\ &\overset{n^+=n^-}{=} G(n^+(-\alpha_1, \dots, -\alpha_I), n^{-\alpha_1, \dots, -\alpha_I}) \\ &= F(-\alpha) \\ &= -F(\alpha), \qquad \text{[by N]} \end{aligned}$$
$$\Rightarrow F(\alpha) = 0$$
    - $n^+(\alpha) > n^-(\alpha) \;\Rightarrow\; F(\alpha) = 1$. Suppose wlog that
$$\alpha = \left(\underbrace{+1, \dots, +1}_{H \text{ individuals}}, \dots, \underbrace{-1, \dots, -1}_{J \text{ individuals}}\right), \qquad H > J$$
$$\alpha' = \left(\underbrace{+1, \dots, +1}_{J}, \dots, \underbrace{-1, \dots, -1}_{J}\right)$$
      $F(\alpha') = 0$ by the previous result. Then PR implies that $F(\alpha) = 1$. Similarly, $n^-(\alpha) > n^+(\alpha) \;\Rightarrow\; F(\alpha) = -1$.
      Therefore, $F(\alpha)$ has to be simple majority rule. ∎

- Example. Pairwise Comparison of Alternatives
  - 3 individuals

$\succcurlyeq_1$	$\succcurlyeq_2$	$\succcurlyeq_3$
x	y	z
y	z	x
z	x	y

  The social preferences (according to simple majority rule):
$$x \succcurlyeq_s y, \qquad y \succcurlyeq_s z, \qquad z \succcurlyeq_s y$$
  But this is not transitive. This is known as the "Condorcet paradox"
    - If we want to avoid this result by ruling out the profiles that lead to this outcome, we lose UD.
    - So when we introduce more than 2 alternatives, we lose May's result completely.

<u>The general case of more than 2 alternatives</u>

- Arrow's Impossibility Theorem.
  - Notations. Recall
    - The social welfare function (SWF) is
      $$F : A \to \mathcal{R}$$
      where $A \subseteq \mathcal{R}^F$ is the set of all possible rational preference relations over X
      $$\succcurlyeq_s = F(\succcurlyeq_1, \dots, \succcurlyeq_I)$$

  - Properties:
    - UD: $\mathcal{A} = \mathcal{R}^I$
    - P: $\forall i : s \succ_i y \;\Rightarrow\; x \succ_s y$
    - IIA (independence of irrelevant alternatives): social ranking of any two alternatives depend <u>only</u> on how individuals rank these two alternatives.
      For any $\{x, y\} \subset X$, $\{\succcurlyeq, \succcurlyeq'\} \subseteq \mathcal{R}^I$, if
      $$x \succcurlyeq_i y \;\Leftrightarrow\; x \succcurlyeq_i' y \;\wedge\; y \succcurlyeq_i x \;\Leftrightarrow\; y \succcurlyeq_i' x$$
      We have
      $$x \succcurlyeq_s y \;\Leftrightarrow\; x \succcurlyeq_s' y \;\wedge\; y \succcurlyeq_s x \;\Leftrightarrow\; y \succcurlyeq_s' x$$
      - Example. Borda rule.

$\succcurlyeq_1$	$\succcurlyeq_2$	$\succcurlyeq_3$	$\succcurlyeq_s$
x	y	z	x
w	z	x	z
y	x	w	y
z	w	y	w

$r_i(x) = n$ where n is the rank in i's preference ordering

$$x: \quad 1 + 3 + 2 = 6$$
$$y: \quad 3 + 1 + 4 = 8$$
$$w: \quad 2 + 4 + 3 = 9$$
$$z: \quad 4 + 2 + 1 = 7$$

If we switch ranking between irrelevant alternatives w, z

$\succcurlyeq_1$	$\succcurlyeq_2$	$\succcurlyeq_3$	$\succcurlyeq_s$
x	y	w	w
y	w	z	y
w	z	x	x
z	x	y	z

$$x: \quad 1 + 4 + 3 = 8$$
$$y: \quad 2 + 1 + 4 = 7$$
$$w: \quad 3 + 2 + 1 = 6$$
$$z: \quad 4 + 3 + 2 = 9$$

This contradicts IIA. Suppose the only irrelevant alternative is w, the outcome also contradict IIA.

- **Theorem (Arrow).** Suppose $|X| \geq 3$. If the social welfare functional F satisfies UD, P, and IIA, the F is dictatorial, i.e. there exists an individual $i \in I$ such that $\forall x \in X$ and $(\succcurlyeq_1, \dots, \succcurlyeq_I) \in \mathcal{R}^I$, we have $x \succ_i y \;\Rightarrow\; x \succ_s y$.

- Social Choice Theorem.
  - *Definition.* a ***social choice function*** (***SCF***) is a function $f : \mathcal{A} \rightarrow X$ that assigns <u>one</u> alternative $x \in X$ to all profiles of individual preference orderings $\succcurlyeq = (\succcurlyeq_1, \dots, \succcurlyeq_I)$ in the admissible domain $\mathcal{A} \subseteq \mathcal{R}^I$.
    - UD: $\mathcal{A} = \mathcal{R}^I$
    - P: $\forall x, y \in X, \forall \succcurlyeq \in \mathcal{A}, \forall i : x \succ_i y \;\Rightarrow\; f(\succcurlyeq) \neq y$
    - M: monotonicity. Suppose $f(\succcurlyeq) = x$. If, $\forall i \in I, \forall y \neq x \in X$ the profile $\succcurlyeq'$ is such that $x \succcurlyeq_i y \;\Rightarrow\; x \succcurlyeq_i' y$, then $f(\succcurlyeq') = x$.

  - **Theorem.** Suppose $|X| > 3$. If the social choice function f satisfies UD, P, and M, then f is dictatorial, i.e. there exists an individual $i \in I$, such that

$$\forall x \in X, \forall(\succcurlyeq_1, \dots, \succcurlyeq_I) \in \mathcal{R}^I : f(\succcurlyeq_1, \dots, \succcurlyeq_I) \in \arg\max\{\succcurlyeq_i |x \in X\}$$

Proof of Arrow's Impossibility Theorem

- Refer to class handout.

Majority Voting

- A way to get around the Arrow's impossibility theorem is to drop the requirement of universal domain (UD), in light of the consideration that not all preferences are likely to occur.
  - ***Restricted Domain***: majority rule the median voter.

- *Definition.* A preference profile $\succcurlyeq \in \mathcal{R}^I$ is ***single-peaked*** if $\forall i \in \{1, \dots, I\}, \exists x_i^* \in X$
  (a) $x_i^* \succ_i y, \ \forall y \in X \setminus \{x_i^*\}$
  (b) $x_i^* \geq z > y \ \Rightarrow \ z \succ_i y$ and $y > z \geq x_i^* \ \Rightarrow \ z \succ_i y$, where "$\geq$" is a linear order on the set of alternatives X.
  - Note that to establish single-peakedness, we only need to find ONE linear order for which (a) and (b) holds. In other words, to reject single-peakedness, we need to check ALL possible linear orders.
  - Note that "single-peakedness" is a statement about the entire profile of individual preferences. This restricts the preference profile to be a specific subset of all the possible preference profiles.

- *Definition.* The individual $m \in I$ is the ***median voter*** or ***median agent*** for the single-peaked preference profile $\succcurlyeq \in \mathcal{R}^I$ if
$$\#\{i \in I | x_i^* \geq x_m^*\} \geq \frac{I}{2} \ \wedge \ \#\{i \in I | x_i^* \leq x_m^*\} \geq \frac{I}{2}$$
  - Note, if I is odd, then m is unique.

- **Proposition.** If the preference profile $\succcurlyeq \in \mathcal{R}^I$ is single-peaked, then x_m^* cannot be defeated by any other alternative in pairwise majority vote:
$$x_m^* \succcurlyeq_M y, \qquad \forall y \in X \setminus \{x_m^*\}$$
where $\succcurlyeq_M$ is simple majority rule social preference ordering. Hence, a Condorcet winner exists and coincides with the media agent's bliss point.
  - *Proof.* By inspection.

- **Theorem (Median Voter Theorem I).** Suppose I is odd, and the strict preference profile $\succ_i \in \mathcal{R}^I$ is single-peaked, then,
$$\forall \{x, y\} \in X : x \succ_m y \ \Rightarrow \ x \succ_M y$$
where m is the median voter. Hence, the social preferences generated by pairwise majority rule are complete and transitive.

- Comparison to institutions of voting in practice.
  - Suppose we define "***majority voting***" as follows:
    - Direct democracy: individual vote directly on X
    - Sincere voting: voters vote for the alternative they actually prefer
    - Open agenda: voting takes place over pairs of alternatives, and the winner in one round is pitched against another alternative in the next round, and the set of alternatives includes all policies.

  Then, we have the following
    - **Corollary 1.** Under the conditions of the MVT I, the median voter's bliss point x_m^* is

the unique equilibrium policy (stable point) under majority voting.

- Can consider this as a positive departure from the normative analysis of aggregating individual preferences. Here, instead of thinking this as a response to Arrow's impossibility theorem by dropping UD (which is quite valid), we think of this as what will actually happen in a direct democracy.

➢ Suppose we define "***political competition***" as follows [Downs (1954)]:
- 2 political parties, $j \in \{A, B\}$, choose platforms $x_j \in X$
- Objective of each party is to maximize the number of votes
- The platform that gets at least half the votes wins, and there is a coin-throw if there is a tie.

Then the following is true:
- **Corollary 2.** Under the conditions of MVT I, the game of political competition has a unique NE in which both parties propose the media voter's bliss point $x_j = x_m^*,\ \forall j$.

❖ Setting the MVT to work (examples).

➢ Redistributive Taxation I.
- $u_i = c_i$, where c is composite consumption commodity
- Policy $(t, g) \in \mathbb{R}^2$ where
  - $t \in [0,1]$ is the proportional income tax
  - g is the lump-sum transfer
- Government budget constraint is a Laffer curve:

$$\underbrace{gI}_{\text{expenditure}} = \left(t - \frac{1}{2}t^2\right)\sum_i y_i$$

  - Note that efficiency dictates that optimal tax rate should be zero.
- Indirect utility

$$v(t, g; y_i) = y_i(1-t) + g$$
$$\Rightarrow\ v(t; y_i) = y_i(1-t) + \left(t - \frac{1}{2}t^2\right)\frac{\sum_i y_i}{I}$$

The first-order condition is

$$\frac{\partial v}{\partial t} = -y_i + (1-t)\frac{\sum_i y_i}{I} = 0$$
$$\Rightarrow\ 1 - t = \frac{y_i}{\sum_i y_i / I}\ \Leftrightarrow\ t^i = 1 - \frac{y_i}{\sum_i y_i / I}$$
$$\Rightarrow\ t^i = \begin{cases} 0 & \text{if } \forall i : y_i \geq \bar{y} \\ 1 - \dfrac{y_i}{\bar{y}} & \text{if } \forall i : y_i < \bar{y} \end{cases}$$

Note that the second derivative is

$$\frac{\partial^2 v}{\partial t^2} < 0$$

That is, the indirect utility is strictly concave, i.e. single-peaked.

Single-peakedness and the fact that the most preferred tax rates are monotone in income implies that the median voter is the person with a median income individual. Therefore, we can invoke the MVT I to conclude that there is a unique equilibrium tax rate either under majority voting or political competition is the median voter's bliss tax rate:

$$t^m = \begin{cases} 0 & \text{if } y^m \geq \bar{y} \\ 1 - \dfrac{y^m}{\bar{y}} & \text{if } y^m < \bar{y} \end{cases}$$

- This says that tax rate is positive if the median income is below average income, which coincides with what we observe in reality.

- Redistributive Taxation II. Same as above except that g_i can be any amount as long as $\sum_i g_i = \left(t - \frac{1}{2}t^2\right)\sum_i y_i$. The indirect utility can be written as

$$v(t, g_i; y_i) = (1 - t)y_i + g_i$$

  - Everybody's bliss point is to tax all the individual's income an redistribute the whole revenue to themselves.
  - This shows that if voting happens in more than one dimensions, we're back to the Condorcet paradox.

Median Voter (Cont'd)

- Examples of MVT (cont'd)
  - Redistributive Taxation III. (same as example 1)
    - $u_i = u(c_i, \ell_i; \theta_i)$, $u_c > 0$, $u_\ell < 0$ (ℓ is labor supply), θ is productivity/wage
    - $c_i = (1-t)\theta_i \ell_i + g$
    - $\theta_i \sim f(\theta_i)$ is the distribution of productivity
    - Government budget constraint:
      - Average labor income: $L(t,g) = \sum_i f(\theta_i)\theta_i \ell_i(t,g;\theta_i)$
      - Per capita budget constraint: $g = tL(t,g)$
    - Indirect utility function as

$$v(t,g;\theta_i) = \max_{\ell_i} u_i\big((1-t)\theta_i\ell_i + g, \ell_i; \theta_i\big)$$

Using the envelope theorem:

$$v_t = u_c\big(-\theta_i \ell_i^*(t,g;\theta_i)\big)$$
$$v_g = u_c$$

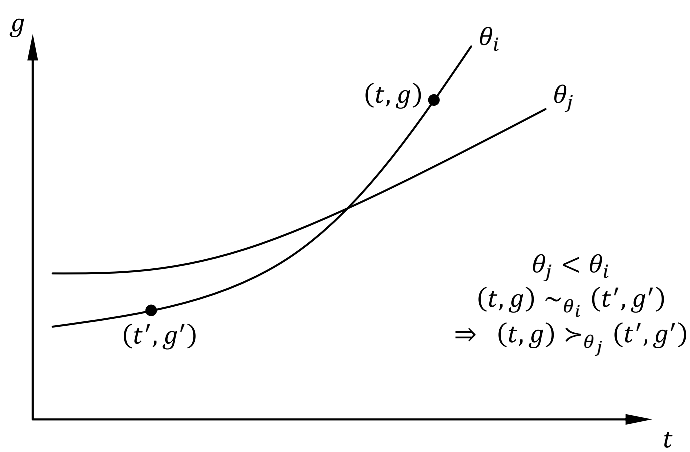
Note that there is NO concavity for v.

  - *Definition.* A preference profile $\succcurlyeq \in \mathcal{R}^I$ satisfies the ***single-crossing property*** (***SC***), if there exists a linear order $\geq$ on X and an order on the agents $\{1,\dots,I\}$ such that

$$\forall x > y, \forall j > i : \big(x \succcurlyeq_i y \;\Rightarrow\; x \succcurlyeq_j y\big) \;\wedge\; \big(x \succ_i y \;\Rightarrow\; x \succ_j y\big)$$

The profile satisfies ***strict single crossing*** (***SSC***) if

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



  - Single Crossing Property as defined in Milgrom and Shannon (1994):
    Let X be a lattice[1], T be a partially ordered set[2], and $f : X \times T \to \mathbb{R}$. Then f satisfies the ***single crossing property*** in $(x;t)$ if

[1] A set X is a ***lattice*** if for every pair $x, y \in X$, the join (or supremum) $x \vee y$ and the meet (or infimum) $x \wedge y$ do exist as elements of X. In other words, a lattice is a partially ordered set in which any two elements have a supremum and an infimum.

[2] A partially ordered set X is one with a binary relation, $\geq$, that is *reflexive* ($\forall x \in X : x \geq x$), *antisymmetric* ($(x \geq y \;\wedge\; y \geq x) \;\Rightarrow\; x = y$), and *transitive*.

$$\forall x' > x'', \forall t' > t'': \begin{cases} f(x', t'') > f(x'', t'') \;\Rightarrow\; f(x', t') > f(x'', t') \\ \text{and} \\ f(x', t'') \geq f(x'', t'') \;\Rightarrow\; f(x', t') \geq f(x'', t') \end{cases}$$

If $f(x', t'') \geq f(x'', t'') \;\Rightarrow\; f(x', t') > f(x'', t')$ for every $t' > t''$, then f satisfies ***strict single crossing property*** in $(x; t)$.

- **Theorem (Median Voter Theorem II).** Suppose I is odd and preferences satisfy (SSC), then

$$\forall \{x, y\} \in X : x \succcurlyeq_m y \;\Leftrightarrow\; x \succcurlyeq_M y$$

where $m = (I+1)/2$. Hence, social preference order generated by pairwise majority rule is complete and transitive.

  - (back to example).

$$\underbrace{\sigma(t, g; \theta_i)}_{MRS_{t,g}} = -\frac{v_t}{v_g} = \frac{u_c \theta_i \ell_i^*(t, g; \theta_i)}{u_c} = \underbrace{\theta_i \ell_i^*(t, g; \theta_i)}_{\text{after tax income of } i}$$

So, $SSC \;\Rightarrow\; \theta_i \ell_i^*$ is increasing in $\theta_i \;\Rightarrow$ most preferred tax rate by θ_m is equilibrium policy.

❖ *Definition.* ***Spence-Mirrlees Condition*** (***SMC***). Let $X \subseteq \mathbb{R}^2$ and $u : X \times I \to \mathbb{R}$ with $u_y > 0$. Then, $u(\cdot)$ satisfies the SMC on X if for all $x \in \text{int}(X)$ and $y \in \mathbb{R}$

$$\sigma(x, y, i) = -\frac{u_x(x, y, i)}{u_y(x, y, i)}$$

is (strictly) increasing in i.

- Note.
  - $u(\cdot)$ satisfies (S)SMC if and only if the preference profile it represents satisfies (S)SC
  - Let $y = f(x)$, the function $u(x, y; i)$ satisfies the (S)SMC if and only if $u(x, f(x), i)$ satisfies (S)SC for all functions $f(x)$.

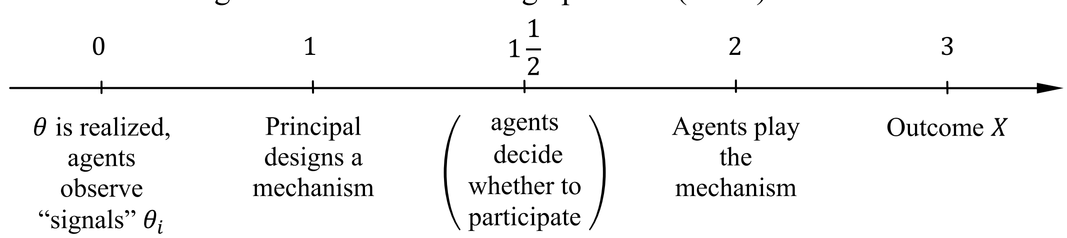
Mechanism Design (Implement Theory)

- Introduction. In the previous section, we looked at ways to aggregate individual preferences into a social preference. However, this relies on the assumption that individual preferences are publicly known. In reality, this is not true. In mechanism design, we look at situations where individual preference profiles are not publicly observable. So if we have a social choice function, we use mechanism design to implement it; or if we can't implement it, mechanism design can tell us how far we can go with it (second best).

- The mechanism design problem – general framework.
  - Agents: $i = 1, \dots, I$
  - Principle (social planner, mechanism designer), may be interpreted as
    - Imaginary player (representing society of agents) $\rightarrow$ social choice problem
    - Real player (government, employer, etc.) $\rightarrow$ principal agent problem
  - Setting: principal designs a mechanism (i.e. a contract) to implement a particular social choice function,
$$f : \mathcal{R}^I \rightarrow X$$
$$(\succcurlyeq_1, \dots, \succcurlyeq_I) \mapsto x$$
where $(\succcurlyeq_1, \dots, \succcurlyeq_I)$ may be private information.
  - How to model this?
    - Each agent i observes a parameter $\theta_i \in \Theta_I$ that determines his/her preferences $\succcurlyeq_i(\theta_i) \in \mathcal{R}_i$. Let SOW = state of the world, which is captured by a vector
$$\theta = (\theta_1, \dots, \theta_I) \in \Theta = \Theta_1 \times \dots \times \Theta_I$$
    - Assume $\succcurlyeq_i(\theta_i)$ can be represented by a VNM utility function $u_i(x, \theta_i)$.

  - *Definition.* A ***social choice function***, $f : \Theta \rightarrow X$ assigns an outcome $x \in X$ to each possible $\theta \in \Theta$.

  - Time structure of a generic mechanism design problem (MDP):



  - *Definition.* A ***mechanism*** $\Gamma = (S_1, \dots, S_I; g(\cdot))$ is a collection of strategy sets $(S_1, \dots, S_I)$, and an outcome function $g : S_1 \times \dots \times S_I \rightarrow X$. A mechanism (contract) is a game form.

  - *Definition.* The mechanism Γ (fully) implements a social choice function $f(\theta)$, if the (unique) equilibrium outcome of Γ in the state of the world θ is $f(\theta)$, i.e.
$$g(s_1^*(\theta_1), \dots, s_I^*(\theta_I)) = f(\theta), \qquad \forall \theta \in \Theta.$$

  - Remarks.
    - θ_i = "type" of agent i, representing characteristics of i, or preferences of i, or private

information about the world.

- Distinguish two environments:
  - θ_i = private information of agent i $\rightarrow$ environment with <u>asymmetric information</u>.
  - θ_i = observed by all agents, but not by the principal/outsider (but this does not mean that θ_i is public information, which is *verifiable*) $\rightarrow$ environment with <u>complete information</u>.

- What is meant by "equilibrium" depends on the solution concept that is used.
  - Environment with asymmetric information
    - Bayesian Nash equilibrium
    - Dominant strategy equilibrium
  - Environment with complete information
    - Nash equilibrium / Subgame Perfect equilibrium
    - Dominant strategy equilibrium
  - Note that since we're *designing* the game, so we can *choose* games that have dominant strategies, e.g. "dominant strategy implementation". This however will restrict the set of possible games that we can choose from.
- What is meant by "equilibrium outcome"
  - Full implementation does not require unique equilibrium (about strategies), only unique *equilibrium outcome*.

- Examples.
  - Public project (e.g. building a bridge).
    - $y \in \{0,1\}$ project decision at cost $c > 0$
    - Outcome $x \in X = \{(y, t_1, \dots, t_I) : y \in \{0,1\},\ \ t_i \in \mathbb{R},\ \ \sum_i t_i \geq cy\}$
    - Individuals: $1, \dots, I$, with
$$u_i(x, \theta_i) = \theta_i y - t_i$$
    - $SOW: \theta = (\theta_1, \dots, \theta_I)$

Social choice function:

- $f(\theta)$ is Pareto efficient if
$$y^*(\theta) = 1 \iff \sum_i \theta_i \geq c$$
- $P(\theta)$ is equal contributions:
$$t_i(\theta) = \frac{c}{I} y(\theta)$$

But this is not implementable. Suppose
$$\forall i \in I \setminus \{1\} : \theta_i = \bar{\theta}_i$$
$$c > \sum_{i \neq 1} \bar{\theta}_i > \frac{c(I-1)}{I}$$
Suppose further that
$$\theta_1 = c - \sum_{i \neq 1} \bar{\theta}_i + \epsilon, \qquad \epsilon > 0$$
If 1 tells the truth,
$$y = 1$$

$$u_1(x,\theta) = \theta_1 - \frac{c}{I} = c - \sum_{i\neq 1} \bar{\theta}_i + \epsilon - \frac{c}{I} = \frac{c(I-1)}{I} - \sum_{i\neq 1} \bar{\theta}_i + \epsilon$$

If 1 lies and reports $\theta_1 = 0$

$$u_1(x,\theta) = 0$$

Then, for small ϵ, lying is better.

- Auction.
  - Principal is the owner of an indivisible object (zero valuation of the good)
  - Two agents, $i = 1,2$
  - Outcome $X = \{(y_1, y_2, t_1, t_2) : y_i \in \{0,1\},\ \sum_i y = 1,\ \ t_i \in \mathbb{R}_0^+\}$
  - Utility: $u_i(x,\theta) = \theta_i y_i - t_i$, where $\theta_i \sim U_{[0,1]}$ is private information.

Example of Mechanism Design

- Auction (cont'd)
  - Two agents: $i = 1,2$
  - One principal who owns the object
  - Outcome:

$$X = \left\{ (y_1, y_2, t_1, t_2) \middle| y_i \in \{0,1\},\ \ y_1 + y_2 = 1,\ \ t_i \in \mathbb{R},\ \ \sum_i t_i \geq 0 \right\}$$

  - Utilities:

$$u_i(x, \theta_i) = \theta_i y_i - t_i$$
$$\theta_i \overset{iid}{\sim} U_{[0,1]}$$

    - Distribution of θ_i is public information, but the individual realization of θ_i is private information
  - Consider the social choice function (SCF):

$$t_i = \theta_i y_i$$
$$y_1 = \begin{cases} 1 & \text{iff } \theta_1 \geq \theta_2 \\ 0 & \text{otherwise} \end{cases}$$

    Note that for efficiency, only the second equation matters; t_i is irrelevant.

  - Is this implementable as a BNE?
    - Suppose $\hat{\theta}_2(\theta_2) = \theta_2$, i.e. $i = 2$ announces truthfully. Then $i = 1$ maximizes

$$\max_{\hat{\theta}_1} E[u_1|\theta_1] = \max_{\hat{\theta}_1} \Pr\{\theta_2 \leq \hat{\theta}_1\} [\theta_1 - \hat{\theta}_1]$$
$$= \max_{\hat{\theta}_1} \hat{\theta}_1 (\theta_1 - \hat{\theta}_1)$$

      Optimal solution: $\hat{\theta}_1 = \frac{1}{2}\theta_1$. There is no NE in which agents announce truthfully.

- First Price Sealed Bid Auction.
  - Two bidders: $i = 1,2$
  - Payoff: $y_i = 1$ if i gets the good

$$u_i = (\theta_i - b_i) y_i, \qquad b_i = \text{bid of } i$$
$$y_i = \begin{cases} 1 & \text{iff } b_i > b_j \\ \text{throw a dice} & \text{if } b_i = b_j \end{cases}$$
$$\theta_i \overset{iid}{\sim} U_{[0,1]}$$

  - To solve the game, guess that $b_i(\theta_i) = \alpha_i \theta_i$ where $\alpha_i \in [0,1]$. Then,

$$\max_{b_1 \in [0, \alpha_2]} (\theta_1 - b_1) \underbrace{\Pr\{b_1 > \alpha_2 \theta_2\}}_{= b_1/\alpha_2}$$

    The solution is

$$b_1 = \begin{cases} \frac{1}{2}\theta_1 & \text{if } \frac{1}{2}\theta_1 \leq \alpha_2 \\ \alpha_2 & \text{otherwise} \end{cases}$$

  - What is the SCF that is implemented by this auction?

$$y_1(\theta) = 1 \iff \theta_1 > \theta_2, \qquad \text{[efficient!!]}$$

$$t_1(\theta) = \frac{1}{2}\theta_1 y_1(\theta)$$
$$t_2(\theta) = \frac{1}{2}\theta_2 y_2(\theta)$$

- Note: this is the exact same SCF as the one implemented by the direct mechanism when agents announce valuations and pay the announced valuations.

Example of Mechanism Design (cont'd)

- We looked at two auctions last time:
  - Direct auction: bidders bid half of their valuation
  - First-price sealed bid auction: bidders also bid half of their valuation

- Second price (Vickery) Auction – [strategically equivalent to oral ascending auction]
  - Buyers: $i = 1,2$
  - Submit bids: $b_i \geq 0$
  - Buyers with the highest bid gets the good, but pays only the second highest bid

  - **Claim.** The strategy $b_i = \theta_i$ is a weakly dominant strategy.
    - Bidders bid truthfully because bidding their true valuations only affect their probability of getting the object, not the price they pay.
    - *Proof.* The payoff of bidder i is

$$u_i = \theta_i y_i + money$$

      - Suppose i bids $b_i = \theta_i$.

$$u_i = \begin{cases} \theta_i - b_j \geq 0 & \text{if } \theta_i \geq b_j \\ 0 & \text{otherwise} \end{cases}$$

      - Suppose i bids $b_i > \theta_i$.

$$u_i = \begin{cases} \theta_i - b_j < 0 & \text{if } b_i > b_j \geq \theta_i \\ \theta_i - b_j \geq 0 & \text{if } b_i > \theta_i \geq b_j \\ 0 & \text{otherwise} \end{cases}$$

      - Suppose i bids $b_i < \theta_i$.

$$u_i = \begin{cases} \theta_i - b_j > 0 & \text{if } b_i > b_j \\ 0 & \text{otherwise} \end{cases}$$

      Case 1 dominates Case 2 when $b_i \geq b_j > 0$, and Case 1 dominates Case 3 when $\theta_i > b_i$ but $b_i < b_j$.

    - The social choice function that is implemented with this auction is

$$y_i(\theta) = \begin{cases} 1 & \text{if } \theta_i \geq \theta_j \\ 0 & \text{otherwise} \end{cases}$$

$$t_i(\theta) = \theta_j y_i(\theta)$$

      - Note that this is a different SCF implemented by the first-price sealed bid auction:

$$y_i(\theta) = \begin{cases} 1 & \text{if } \theta_i \geq \theta_j \\ 0 & \text{otherwise} \end{cases}$$

$$t_i(\theta) = \frac{1}{2}\theta_i y_i(\theta_i)$$

The Revelation Principle

- *Definition.* A ***direct revelation mechanism*** is a mechanism $\Gamma_D = (S_1, \dots, S_I; g(\cdot))$ with
$$S_i = \Theta_i,\ \ \forall i \in I \quad \text{and} \quad g(\theta) = f(\theta).$$

- *Definition.* The social choice function $f(\theta)$ is ***truthfully implementable*** (or ***incentive compatible***) if the direct revelation mechanism Γ_D has an equilibrium in which
$$s_i^*(\theta_i) = \theta_i, \qquad \forall i \in I, \forall \theta_i \in \Theta_i$$
  - i.e. if telling the truth is an equilibrium.

- The ***Revelation Principle*** (informal definition).
Suppose the SCF $f(\theta)$ is (fully) implementable. Then, $f(\theta)$ is also truthfully implementable, i.e. there exists the direct mechanism $\Gamma_D = (\Theta_1, \dots, \Theta_I, f(\theta))$ has an equilibrium in which everybody tells the truth.

Dominant Strategy Implementation

- ***Revelation Principle (in dominant strategies)***. Suppose $f(\theta)$ that is (fully) implementable in dominant strategies, then $f(\theta)$ is also truthfully implementable in dominant strategies; i.e. the direct mechanism $\Gamma_D = \{(\Theta_1, \ldots, \Theta_I); f(\theta)\}$ is such that

$$\forall \theta_i \in \Theta_i : u_i(f(\theta_i, \theta_{-i}); \theta_i) \geq u_i\big(f\big(\tilde{\theta}_i, \theta_{-i}\big); \theta_i\big)$$

for all $\theta_{-i} \in \Theta_{-i}$ and all $\tilde{\theta}_i \in \Theta_i$.

  - *Proof.* If $f(\theta)$ is implementable in dominant strategies, we know there exists an indirect mechanism Γ such that

$$u_i\big((s_i^*(\theta), s_{-i}); \theta_i\big) \geq u_i(\tilde{s}_i, s_{-i}; \theta_i), \qquad \forall i,\ \forall \tilde{s}_i,\ \forall s_{-i},\ \forall \theta_i$$

and

$$g\big(s_i^*(\theta_i), s_{-i}(\theta_{-i})\big) = f(\theta)$$

This must be true, in particular, for

$$\tilde{s}_i = s_i^*\big(\tilde{\theta}_i\big)$$
$$s_{-i} = s_{-i}^*(\theta_{-i})$$

Then,

$$u_i(s_i^*(\theta_i), s_{-i}^*(\theta_{-i}); \theta_i) \geq u_i\big(s_i^*\big(\tilde{\theta}_i^*\big), s_{-i}^*(\theta_{-i}); \theta_i\big), \qquad \forall i,\ \forall \theta_i,\ \forall \tilde{\theta}_i,\ \forall \theta_{-i}$$

But we know

$$g\big(s_i^*(\theta_i), s_{-i}^*(\theta_{-i})\big) = f(\theta), \qquad \forall \theta$$
$$\Rightarrow\ u_i(f(\theta_i, \theta_{-i}); \theta_i) \geq u_i\big(f\big(\tilde{\theta}_i, \theta_{-i}\big); \theta_i\big), \qquad \forall i,\ \forall \theta_i,\ \forall \tilde{\theta}_i,\ \forall \theta_{-i}$$

This completes the proof.

  - Note.
    - This works for other equilibrium concepts as well (obviously)
    - Requires commitment by the principal.

- **Gibbard-Satterthwaite Theorem.** (also, cf. MWG Prop. 21.E.2.)
  - **Lemma 1.** Suppose $\mathcal{R}_i$ only contains strict preferences, and $f(\theta)$ is truthfully implementable (aka "strategy proof"), the $f(\theta)$ is monotonic.
    - This links implementability to the property of SCF.
    - *Proof.* Suppose $f(\theta) = x$, and for all $y \in X$, θ_i' is such that agent i prefers x to y whenever he prefers x to y in θ. We need to show that $f(\theta_i', \theta_{-i}) = x$.
Suppose for contradiction that, when the true state is $\theta = (\theta_i, \theta_{-i})$,

$$f(\theta_i', \theta_{-i}) = y \neq x.$$

Agent i could get y in state $\theta' = (\theta_i', \theta_{-i})$ but doesn't lie (by assumption). So i must rank x above y in state $\theta = (\theta_i, \theta_{-i})$.

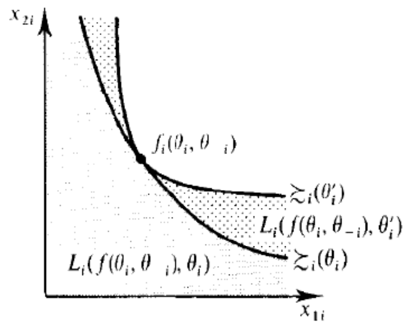
Same reasoning, if the true state is θ', then i could lie and claim θ_i (and get x). But that's not optimal by assumption. So i must rank y above x in θ_i'.

Since no alternatives can be indifferent, need to have

$$f(\theta_i', \theta_{-i}) = (\theta_i, \theta_{-i}) = x$$

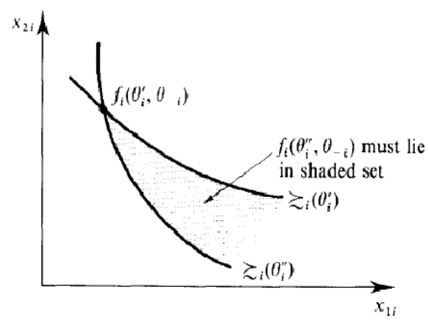
Repeat the argument, one at a time, for all other individuals to get

$$f(\theta_1, \ldots, \theta_I) = f(\theta_1', \ldots, \theta_I')$$



- **Lemma 2.** If $f(\theta)$ is monotonic and onto (i.e. $f(\Theta) = X$), then $f(\theta)$ is efficient.
  - *Proof.* Choose $x \in X$, because f is onto, $f(\theta) = x$ for some θ. By M, $f(\theta') = x$ because
$$u(x, \theta_i') > u(y, \theta_i'), \qquad \forall y \in X$$
Therefore, $f(\theta)$ is efficient.

- **Theorem (Gibbard-Satterthwaite).** Suppose $|X| \geq 3$, and strict preferences only, and $f(\Theta) = X$. Then, the social choice function $f(\theta)$ is truthfully implementable in dominant strategies if and only if $f(\theta)$ is dictatorial, i.e.
$$\exists i, \forall \theta : f(\theta) \in \arg\max u_i(x, \theta_i)$$
  - *Proof.* "⇒" Lemma 1 + Lemma 2 + Theorem III of part 2
"⇐" show for yourself.



❖ Example. Clarke-Groves Mechanism (with quasi-linear preferences)
- Assume
$$X = \left\{ (y, t_1, \dots, t_I) \middle| y \in Y, t_i \in \mathbb{R}, \sum_i t_i \leq 0 \right\}$$
$$u_i(x, \theta_i) = v(y, \theta_i) + t_i$$
- Pareto Optimality:
  - Hold all but one agent's utility to a certain constant level, then maximize the utility of the remaining agent's utility subject to resource constraint.
    - Suppose there are only two agents with quasi-linear utility functions. Then, Pareto optimality is given by
$$\max u_1 = v_1(y) + t_1, \qquad s.t. \begin{cases} u_2 = v_2(y) + t_2 \geq \bar{u} \\ t_1 + t_2 = 0 \end{cases}$$
$$t_1 + t_2 = 0 \ \Rightarrow \ t_2 = -t_1$$
$$\Rightarrow \ u_2 = v_2(y) - t_1 = \bar{u}$$

$$\Rightarrow \; t_1 = v_2(y) - \bar{u}$$

Hence, the objective function becomes

$$\max v_1(y) + v_2(y)$$

Thus, finding Pareto optimality is the same as maximizing the sum of individual utilities. The solution is

$$y^*(\theta) \in \arg\max \sum_i v_i(y, \theta_i)$$

Dominant Strategy Implementation (cont'd)

- Clarke-Groves Mechanism
  - Environment (see last time).
  - **Theorem** (Clarke, Groves, Vickery)**.** There is a social choice function $f(\theta)$ with an efficient outcome

$$y^*(\theta) \in \arg\max \sum_i v_i(y, \theta_i)$$

that can be implemented in dominant strategies.

- *Proof.* Consider the following (direct) mechanism (the "***Groves Mechanism***"): Recall that $\hat{\theta}_i$ is the announced θ_i of agent i (which may or may not be true)

$$y^*(\hat{\theta}) \in \arg\max \sum_i v_i(y, \hat{\theta}_i), \qquad \forall \hat{\theta} \in \Theta$$

$$t_i(\hat{\theta}) = \underbrace{\sum_{j \neq i} v_j\big(y(\hat{\theta}_i \hat{\theta}_{-i}); \hat{\theta}_j\big)}_{\substack{\text{externality that } i \\ \text{generates for all other individuals} \\ \text{given their announced } \hat{\theta}_j\text{'s}}} + h_i(\hat{\theta}_{-i})$$

where $h_i(\cdot)$ is some arbitrary function of $\hat{\theta}_{-i}$. Then,

$$u_i(y, t, \theta_i) = v_i(y, \theta_i) + t_i$$

Given this mechanism, agents have a dominant strategy to tell the truth because

$$u_i(y, t, \hat{\theta}_i = \theta_i) = v_i\big(y^*(\theta_i, \hat{\theta}_{-i}), \theta_i\big) + \sum_{j \neq i} v_j\big(y(\theta_i, \hat{\theta}_{-i}), \hat{\theta}_j\big) + h_i(\hat{\theta}_{-i})$$

$$= \sum_i v_i\big(y^*(\theta_i, \hat{\theta}_{-i}), \theta_i, \hat{\theta}_{-i}\big) + constant$$

$$\overset{\text{by def of } y^*}{\geq} \sum_i v_i\big(y^*(\theta_i', \hat{\theta}_{-i}), \theta_i, \hat{\theta}_{-i}\big) + constant$$

for all $\hat{\theta}_{-i}$, all $\theta_i' \neq \theta_i$, and all i.

- The idea is that
  - The outcome of the mechanism is such that y^* is efficient if truth-telling is equilibrium
  - Transfer to each agent comes in two parts:
    (a) Externality given announcements of others ensures truth-telling will maximize surplus (even if others lie!)
    (b) Constant that only depends on what others announce

- Notes:
  - Special case of the Groves mechanism is the ***Clarke*** (or ***pivotal***) ***mechanism***:

$$h_i(\hat{\theta}_{-i}) = -\sum_{j \neq i} v_j\big(y_{-i}^*(\hat{\theta}_{-i}), \hat{\theta}_j\big)$$

where $y_{-i}^*(\hat{\theta}_{-i}) \in \arg\max \sum_{j \neq i} v_j(y, \hat{\theta}_j)$.

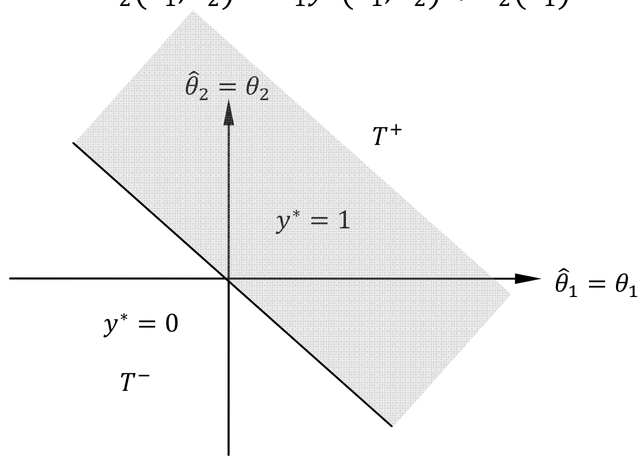
- So you pay only if you are pivotal, and your payment equals the net externality imposed on others.

- Special case of Clarke mechanism is the ***Vickery Auction*** (i.e. second price seal-bid auction / English auction).
- Reverse of the theorem also holds: If the space of the utility functions is sufficiently rich, then every incentive compatible mechanism is a Groves mechanism.
- However, the outcome of the Groves mechanism is not ex post efficient because the budget is not balanced in general.

Groves Mechanism

- **Claim.** Groves Mechanism is not balanced in general.
  - **Corollary.** There does not exist a mechanism that implements the efficient allocation in dominant strategy in general.

  - *Proof* (by example). Suppose $I = 2$, $y = \{0,1\}$ (costless public project) with
$$u_i(y, \theta_i) = \theta_i y + t_i, \qquad \theta_i \in \Theta_i = \mathbb{R}$$
The efficient allocation is
$$y^*(\theta_1, \theta_2) = 1 \quad \text{iff } \theta_1 + \theta_2 \geq 0$$
$$t_1(\theta_1, \theta_2) + t_2(\theta_1, \theta_2) = 0, \qquad \forall \theta_1, \theta_2$$
Groves Mechanism is a direct mechanism, so it takes the announced type, $\hat{\theta}_1, \hat{\theta}_2$, as arguments:
$$y^*(\hat{\theta}_1, \hat{\theta}_2) = 1 \quad \text{iff } \hat{\theta}_1 + \hat{\theta}_2 \geq 0$$
$$t_1(\hat{\theta}_1, \hat{\theta}_2) = \hat{\theta}_2 y^*(\hat{\theta}_1, \hat{\theta}_2) + h_1(\hat{\theta}_2)$$
$$t_2(\hat{\theta}_1, \hat{\theta}_2) = \hat{\theta}_1 y^*(\hat{\theta}_1, \hat{\theta}_2) + h_2(\hat{\theta}_1)$$



Show: if the mechanism is balanced on T^+, it will be unbalanced in T^-.
Recall that $\hat{\theta}_i = \theta_i$ in Groves mechanism.
$$(\theta_1, \theta_2) \in T^+ \;\Rightarrow\; y^*(\theta_1, \theta_2) = 1$$
$$t_1(\theta_1, \theta_2) = \theta_2 + h_1(\theta_2)$$
$$t_2(\theta_1, \theta_2) = \theta_1 + h_2(\theta_1)$$
$$t_1(\theta_1, \theta_2) + t_2(\theta_1, \theta_2) = \theta_1 + \theta_2 + h_1(\theta_2) + h_2(\theta_1) = 0$$
$$\Rightarrow\; \theta_1 + h_2(\theta_1) = -\theta_2 - h_1(\theta_2), \qquad \forall (\theta_1, \theta_2)$$
$$\Rightarrow\; (*) \begin{cases} \theta_1 + h_2(\theta_1) = \gamma & \forall \theta_1 \\ \theta_2 + h_1(\theta_2) = -\gamma & \forall \theta_2 \end{cases}$$
Now consider $(\theta_1, \theta_2) \in T^- \;\Rightarrow\; y^*(\theta_1, \theta_2) = 0$
$$\Rightarrow\; \begin{cases} t_1(\theta_1, \theta_2) = h_1(\theta_2) \\ t_2(\theta_1, \theta_2) = h_2(\theta_1) \end{cases}$$
From $(*)$,
$$\Rightarrow\; t_1 + t_2 = -\gamma - \theta_2 + \gamma - \theta_1$$
$$= -(\theta_1 + \theta_2)$$
$$\geq 0$$
The inequality is strict when $\theta_1 + \theta_2 < 0$. So there is a surplus on T^-. Therefore the sum

is unbalanced!

- Notes.
  - Budget balance (ex post efficiency) can be overcome if either
    (a) There is a principal = 3rd party who can break the budget (e.g. second price sealed bid auction)
    (b) There is at least one agent whose preferences are known → give deficit / surplus generated to that "outside"
  - Results do depend on the "richness" of the type space. For instance, $\Theta = \{-2,1\}$ and check.
  - Large number of agents can lead to approximate efficiency.
    - For example, get surplus to one individual, then efficiency loss per agent $\rightarrow 0$.
  - Individual rationality (participation), once types are known, is not necessarily ensured as long as no outside source (3rd party financing deficit) is available.

Bayesian Nash Implementation

- Environment – with incomplete and asymmetric information
  - State of the World: $\theta = (\theta_1, \dots, \theta_I)$, drawn from $\Theta = \Theta_1 \times \cdots \times \Theta_I$ according to some probability density function $\phi(\theta)$
  - Agent's utility function $u_i(x, \theta_i)$ is a VNM utility function
  - Each agent privately observes his own θ_i only, but holds beliefs about $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_I)$ ex ante, those beliefs coincide with $\phi(\theta)$, and are common knowledge.

- Mechanism $\Gamma = \left(S_1, \dots, S_I; g(\cdot)\right)$ where $g : S \to X$ defines, together with the set of agents, agents' utility functions, and density ϕ, a game of incomplete information. The equilibrium concept is the BNE (and refinements thereof).
  - Recall: BNE = NE ex ante or in expectation.

- *Definition.* The mechanism Γ ***implements*** the social choice function $f(\cdot)$ in BNE if there is a BNE of Γ, $\left(s_1^*(\theta_1), \dots, s_I^*(\theta_I)\right)$ such that
$$g\left(s^*(\theta)\right) = f(\theta), \qquad \forall\theta.$$

- *Definition.* The direct mechanism Γ_D in which $s_i(\theta_i) = \hat{\theta}_i(\theta_i)$, is ***truthfully implementable*** if it has a BNE in which
$$\forall i, \forall\theta_i \in \Theta_i, \forall\theta_i' \neq \theta_i : E_{\theta_{-i}}[u_i(f(\theta_i, \theta_{-i}); \theta_i | \theta_i] \geq E_{\theta_{-i}}[u_i(f(\theta_i', \theta_{-i}); \theta_i) | \theta_i]$$
  - Note that $E_{\theta_{-i}}$ already takes into account that everybody tells the truth.

- ***Revelation Principal for BNE***. If $f(\theta)$ is (fully) implementable in BNE, then $f(\theta)$ is truthfully implementable in BNE, i.e. the direct mechanism $\Gamma_D = \left(\theta_1, \dots, \theta_I, f(\theta)\right)$ has $\hat{\theta}_i = \theta_i$ for all θ_i.
  - Note. In general, full truthful implementation in BNE is not guaranteed. In other words, there may be other BNE's in which people lie in the direct mechanism.

- **Revenue Equivalence Theorem.** Assume each of a given number I risk neutral buyers of an object has a privately known signal θ_i that is independently drawn from some interval $\left[\underline{\theta}_i, \overline{\theta}_i\right]$ with positive density $\phi_i(\theta_i) > 0$ everywhere. Then, any auction (mechanism) in which
  (a) the object goes to the buyer with the highest signal
  (b) any buyer with the lowest feasible signal expects no surplus

  yields the exact same expected revenue for the seller.

Proof of Revenue Equivalence Theorem

- **Revenue Equivalence Theorem.** Assume each of a given number I risk neutral buyers of an object has a privately known signal θ_i that is independently drawn from some interval $[\underline{\theta}_i, \overline{\theta}_i]$ with positive density $\phi_i(\theta_i) > 0$ everywhere. Then, any auction (mechanism) in which
  (a) the object goes to the buyer with the highest signal
  (b) any buyer with the lowest feasible signal expects no surplus
  yields the exact same expected revenue for the seller.

  - Note that the theorem applies to both the
    - Private value model: θ_i is the individual private valuations of bidder i
    - Common value model: θ_i signal about a common value of the object being sold, e.g. oil field.

  - *Proof* (for independent private values). θ_i is the valuation of i.
    The revelation principle implies that we can wlog restrict attention to direct revelation mechanisms (DRM) with
    $$u_i = \theta_i y_i - t_i = \theta_i y_i(\theta_i) - t_i(\theta_i)$$
    where
    $$y_i(\theta_i) = \text{probability of receiving the object if valuation is } \theta_i$$
    $$t_i(\theta_i) = \text{transfer to seller if valuation (also the announcement) is } \theta_i$$

    Expected utility of i under the DRM
    $$\begin{aligned} u_i(\theta_i) &= E_{-\theta_i}[\theta_i y_i(\theta_i, \theta_{-i}) - t_i(\theta_i, \theta_{-i})] \\ &= \theta_i \bar{y}_i(\theta_i) - \bar{t}_i(\theta_i) \end{aligned}$$

    Bayesian incentive compatibility requires that $\forall \hat{\theta}_i \neq \theta_i$,
    $$u_i(\theta_i) = \theta_i \bar{y}_i(\theta_i) - \bar{t}_i(\theta_i) \geq \theta_i \bar{y}_i(\hat{\theta}_i) - \bar{t}_i(\hat{\theta}_i), \qquad (*)$$
    $$u_i(\hat{\theta}_i) = \hat{\theta}_i \bar{y}_i(\hat{\theta}_i) - \bar{t}_i(\hat{\theta}_i) \geq \hat{\theta}_i \bar{y}_i(\theta_i) - \bar{t}_i(\theta_i), \qquad (**)$$

    $$(*) \;\Leftrightarrow\; u_i(\theta_i) \geq u_i(\hat{\theta}_i) + (\theta_i - \hat{\theta}_i)\bar{y}_i(\hat{\theta}_i)$$
    $$(**) \;\Leftrightarrow\; u_i(\hat{\theta}_i) \geq u_i(\theta_i) + (\hat{\theta}_i - \theta_i) - \bar{y}_i(\theta_i)$$

    $$\bar{y}_i(\hat{\theta}_i) \overset{(*)}{\geq} \frac{u_i(\hat{\theta}_i) - u_i(\theta_i)}{\hat{\theta}_i - \theta_i} \overset{(**)}{\geq} \bar{y}_i(\theta_i), \qquad \forall \theta_i, \hat{\theta}_i$$
    Let $\hat{\theta}_i \to \theta_i$. Then,
    $$\frac{\partial u_i(\theta_i)}{\partial \theta_i} = \bar{y}_i(\theta_i)$$
    This means that the equilibrium utilities (in <u>any</u> game with the same $\bar{y}_i(\theta_i)$) are fully determined, up to a constant!! Integrate up gives
    $$u_i(\theta_i) = u_i(\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\theta_i} \bar{y}_i(\tau) d\tau$$
    So the expected utility of individual i. Now consider any two auction that have the same $u_i(\underline{\theta}_i)$ and the same probability of receiving the object, $\bar{y}_i(\theta_i)$, for all θ_i and all i. These

two auctions will generate the same $u_i(\theta_i) = \theta_i \bar{y}_i(\theta_i) - \bar{t}_i(\theta_i)$, then

- expected payments $\bar{t}_i(\theta_i)$ must be the same in both auctions; and
- expected revenue of the seller is the same. ■

❖ Remark.

When true state is θ_i

$$\theta_i \bar{y}_i(\theta_i) - \bar{t}_i(\theta_i) \geq \theta_i \bar{y}_i(\hat{\theta}_i) - \bar{t}_i(\hat{\theta}_i)$$

When true state is $\hat{\theta}_i$

$$\hat{\theta}_i \bar{y}_i(\hat{\theta}_i) - \bar{t}_i(\hat{\theta}_i) \geq \hat{\theta}_i \bar{y}_i(\theta_i) - \bar{t}_i(\theta_i)$$

Then,

$$\hat{\theta}_i \left(\bar{y}_i(\hat{\theta}_i) - \bar{y}_i(\theta_i) \right) \geq \bar{t}_i(\theta_i) - \bar{t}_i(\hat{\theta}_i) \geq \theta_i \left(\bar{y}_i(\hat{\theta}_i) - \bar{y}_i(\theta_i) \right)$$

$$\Rightarrow \ (\hat{\theta}_i - \theta_i)\left(\bar{y}_i(\hat{\theta}_i) - \bar{y}_i(\theta_i) \right) \geq 0$$

$$\Rightarrow \ \hat{\theta}_i \geq \theta_i \ \Rightarrow \ \bar{y}_i(\hat{\theta}_i) \geq \bar{y}_i(\theta_i)$$

That is, the probability of winning the object has to be non-decreasing in θ_i.

Bayesian Nash Implementation (cont'd)

❖ **Theorem.** There is a Pareto efficient social choice function $f(\theta)$ that can be truthfully implemented in Bayesian Nash equilibrium (BNE) in the following setting:

- Quasi-linear preferences: $u_i(x, \theta_i) = v_i(y, \theta_i) + t_i$
- θ_i's are drawn independently

- *Proof.* Consider the expected externality mechanism (or d'Apresmont Gerard-Varet mechanism):

$$y^*(\hat{\theta}) \in \arg\max_y \sum_i v_i\big(y(\hat{\theta}), \hat{\theta}_i\big)$$

$$t_i(\hat{\theta}) = \underbrace{E_{\theta_{-i}}\left[\sum_{j \neq i} v_j\big(y^*(\hat{\theta}_i, \theta_{-i}), \theta_j\big)\right]}_{H_i(\theta_i)} + h_i(\theta_{-i})$$

  - Note that in truth telling equilibrium $\hat{\theta}_{-i} = \theta_i$
  - $y^*(\hat{\theta})$ is ex post efficient
  - $H_i(\theta_i)$ depends <u>only</u> on θ_i, because all the θ_{-i}'s are expected out

  - Is it optimal to tell the truth? → Yes.

$$\hat{\theta}_i \in \arg\max E_{\theta_{-i}}\left[v_i\big(y^*(\hat{\theta}_i, \theta_{-i}), \theta_i\big) + \sum_{j \neq i} v_j\big(y^*(\hat{\theta}_i, \theta_{-i}), \theta_j\big) + h_i(\theta_{-i})\right]$$

$$\Rightarrow \; \hat{\theta}_i = \theta_i, \qquad \left[\because \; y^*(\theta) = \arg\max \sum_i v_i\big(y(\hat{\theta}), \hat{\theta}_i\big)\right]$$

  - We can get budget balance. Since $H_i(\hat{\theta}_i)$ does <u>not</u> depend on $\hat{\theta}_{-i}$, so it can be paid for by others without distorting their incentives to reveal the truth. One possibility is

$$h_i(\hat{\theta}_{-i}) = -\frac{1}{I-1}\sum_{j \neq i} H_j(\hat{\theta}_j)$$

$$\Rightarrow \; t_i(\hat{\theta}_i) = H_i(\hat{\theta}_i) - \frac{1}{I-1}\sum_{j \neq i} H_j(\hat{\theta}_j)$$

$$\Rightarrow \; \sum_i t_i(\hat{\theta}_i) = \sum_i H_i(\hat{\theta}_i) - \frac{1}{I-1}\sum_i (I-1) H_i(\hat{\theta}_i) = 0$$

This completes the proof.

  - Notes:
    - There may be other BNE where everybody lies (i.e. not full implementation)
    - The mechanism is still not individually rational (in the *interim* sense, i.e. given people know their own valuations)

- **Theorem (Myerson-Satterthwaite).** Consider a bilateral trade setting where the buyer and the seller are risk neutral. Suppose their valuations θ_S and θ_B are drawn independently from distributions with strictly positive densities over $[\underline{\theta}_S, \bar{\theta}_S]$ and $[\underline{\theta}_B, \bar{\theta}_B]$, and θ_B and θ_S are private information. Then, there does not exist a mechanism that
  (a) implements the efficient allocation in BNE, and
  (b) is individually rational (i.e. gives non-negative expected gains from trade).

- Solomon Example.
  - Two agents, A, B
  - Two states of the world, α, β
  - Environment with complete (but not verifiable) information
  - Outcomes: $X = \{a, b, c, d\}$
  - Social choice function: $f(\alpha) = a$ and $f(\beta) = b$
  - Preferences.
  - Look at direct revelation mechanism
  - "shoot them all" mechanism

Second-Degree Price Discrimination

- Optimal pricing scheme for a monopolist who do not know the preferences of its customer

- Two parties
  - Principal (monopolist)
  - Agent (customer)
- Principal sells a good to the agent
  - Outcome $X = (y, t)$, where y is quantity sold / consumed and t is price
- Preferences

$$u_A = \underbrace{v(y,\theta)}_{\oplus} - \underbrace{t}_{\ominus}, \qquad u_P = t - cy, \qquad c \geq 0$$

- Distribution of θ:

$$\theta \in \{\theta_L, \theta_H\}, \qquad \Pr(\theta = \theta_H) = p$$

- Efficiency:

$$y^* \in \arg\max v(y,\theta) - cy$$

  The FOC is

$$v'(y^*, \theta) = c$$

  t doesn't matter.

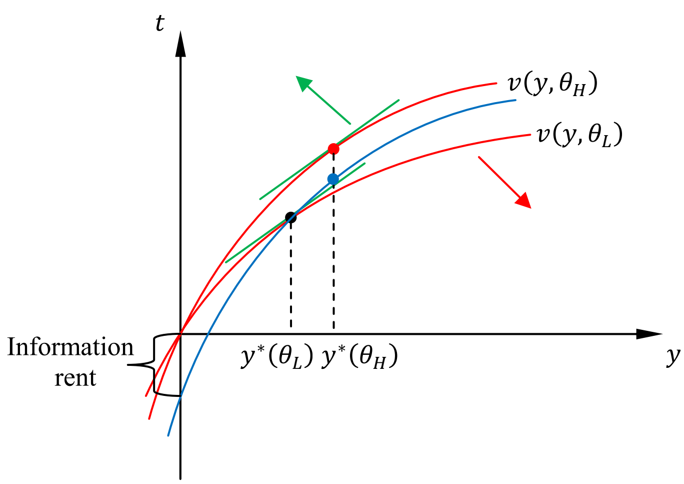
- Profit maximizing for the monopolist (if θ is known)

$$\max_{t,y} t - cy, \qquad s.t. \quad \underbrace{v(y,\theta) - t \geq 0}_{\text{participation constraint}}$$

  This is the Pareto program.

$$y^{FB} = y^*, \qquad t^{FB} = v(y^*, \theta)$$

  This outcome is efficient!
  - Note that the monopoly outcome is efficient because θ is known.



  Assume:

$$v(y, \theta_L) < v(y, \theta_H), \qquad y > 0$$
$$v'(y, \theta_L) < v'(y, \theta_H), \qquad \forall y$$

  - Example: $v(y, \theta) = \theta v(y)$
  - If θ is not known, the first best allocation (i.e. the black and red dots) cannot be

implemented because it is not incentive compatible.

- However, the efficient y can be implemented (the black and blue dots)

- Second best

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

subject to

$$v(y_H,\theta_H) - t_H \geq 0$$
$$v(y_L,\theta_L) - t_H \geq 0$$
$$v(y_H,\theta_H) - t_H \geq v(y_L,\theta_H) - t_L$$
$$v(y_L,\theta_L) - t_L \geq v(y_H,\theta_L) - t_H$$

Both constraints are binding. FOC is

$$v'(y_H,\theta_H) = c \;\; \Rightarrow \;\; y_H = y_H^*$$
$$v'(y_L,\theta_L) > c \;\; \Rightarrow \;\; y_L < y_L^*$$

<u>Moral Hazard Problem</u>

- In a moral hazard problem, the information asymmetry arises <u>after</u> the mechanism (or contract) is designed.
  - In comparison, in the adverse selection problem, the information asymmetry issue is ex ante—i.e. the principal is aware of the information asymmetry when designing the mechanisms

- Examples
  - Employee—employer
  - Lawyer—client
  - Insure—insurance company
  - CEO—stockholders

- Environment:
  - Principal, agent (two parties)
  - Principal owns a technology $F(x; a)$, where
    $$x \in \{x_1, \dots, x_n : x_1 < x_2 < \cdots < x_n\}$$
    is an observable and verifiable (i.e. the court can observe it) outcome.
    $$a \in A$$
    is an unobserved action taken by the agent (e.g. effort, investment decision, attention,…).
    $$F(x; a)$$
    is a probability distribution over x given a, with
    $$f_i(a) = \Pr\{x = x_i | a\} > 0, \qquad \forall a, \ \forall i$$
  - The principal does not want to, or cannot, choose a herself. So she delegates a to the agent.
  - Although a is not observable, the principal can contract indirectly on a through paying a wage $w(x)$. This is the "incentive wage".
  - Utility function for the principal:
    $$u_P(x, a, w) = v(x) - w(x)$$
    - Note that the principal is risk neutral.
  - Utility function for the agent:
    $$u_A(w, a) = u\big(w(x)\big) - g(a), \qquad u' > 0, \qquad u'' \leq 0$$

- First-best solution (directly contract on a as if it is observable)
  - Suppose a is observable / verifiable, then we can directly contract on a. The principal can induce any desired $\hat{a}$ by specifying a wage
    $$w(a) = \begin{cases} \bar{w} & \text{if } a = \hat{a} \\ -\infty & \text{if } a \neq \hat{a} \end{cases}$$
  - What is the best $\hat{a}$ and the best $w(\hat{a}, x)$ for the principal? We can solve for these in two steps:
    1. Suppose the principal wants to induce $a \in A$, what is the optimal (i.e. cost minimizing) wage scheme, $w(x)$, for this particular action $\rightarrow$ wage cost $c(a)$
    2. Which action does the principal want the agent to take, given the corresponding wage cost $c(a)$.

- Step 1: take a as given, let $w_i = w(x_i)$

$$\max_{w_i \in \{w_1,\dots,w_n\}} E(u_P) = \sum_i f_i(a)(v(x_i) - w_i)$$

subject to

$$E(u_A) = \sum_i f_i(a)\big(u(w_i) - g(a)\big) \geq \bar{u}$$

  - Note that this is the Pareto program, so the solution will be efficient.

Use a Lagrangian:

$$\mathcal{L} = \sum_i f_i(a)(v(x_i) - w_i) + \lambda \left[\sum_i f_i(a)\big(u(w_i) - g(a)\big) - \bar{u} \right]$$

The first-order condition is

$$\frac{\partial \mathcal{L}}{\partial w_i} = -f_i(a) + \lambda f_i(a) u'(w_i) = 0 \;\Leftrightarrow\; \lambda u'(w_i) = 1, \qquad \forall i$$

$$\Rightarrow\; w_1 = w_2 = \cdots = w_n$$

The wage is constant (over all outcome). This requires that $u'' < 0$.

  - If the agent is risk averse, the risk neutral principal should insure the agent fully, i.e. should not expose the agent to any income risk.
  - It is costly to pay the agent different wages based on different outcomes, because agents are risk averse while the principal is not.

So in the first-best solution, the principal pays a fixed wage, which will be such that the participation constraint is binding:

$$u(w) - g(a) = \bar{u} \;\Rightarrow\; w = u^{-1}\big(\bar{u} + g(a)\big) = c^{FB}(a)$$

- Step 2: given $c^{FB}(a)$, the principal chooses a so as to

$$\max_{a \in A} E(u_P) = \sum_i f_i(a)[v(x_i)] - c^{FB}(a)$$

  - Implicit assumption: the principal's participation constraint is satisfied.

- Moral hazard: a is not unobservable.
  - If the principal still pays the flat wage, then the agent is going to choose

$$\max_{a \in A} u_A = u(w) - g(a) \;\Rightarrow\; a = \arg\min_{a \in A} g(a)$$

The "least cost action".

  - Solve the second best contract in the same two steps as above.

  - Step 1: cost-minimizing way of inducing some action a

$$\min_{w_i \in \{w_1,\dots,w_n\}} \sum_i f_i(a) w_i$$

subject to

$$\sum_i f_i(a) u(w_i) - g(a) \geq \bar{u}$$

$$\sum_i f_i(a)u(w_i) - g(a) \geq \sum_i f_i(a_j)u(w_i) - g(a_j), \qquad \forall a_j \neq a$$

The constraints are not necessarily convex. To make it convex, use variable transformation: $u_i = u(w_i)$. Then, letting $v(\cdot) = u^{-1}(\cdot)$, the problem becomes

$$\min_{u_i \in \{u_1,\dots,u_n\}} \sum_i f_i(a)v(u_i)$$

subject to

$$\sum_i f_i(a)u_i - g(a) \geq \bar{u}$$

$$\sum_i f_i(a)u_i - g(a) \geq \sum_i f_i(a_j)u_i - g(a_j), \qquad \forall a_j \neq a$$

- Note: participation is binding. Suppose it does not bind, then the principal could always reduce $\hat{u}_i = u_i - \epsilon$ and make herself better off.
- Whenever $a \neq \arg\min_a g(a)$, the incentive compatibility constraint for at least one action $a_j \neq a$ is binding. So the agent is indifferent between at least two actions under the optimal contract.
- Why would the agent being indifferent between taking two actions take the action that the principal wants him to take? Because otherwise there would be no best response for the principal when there is a positive probability that the agent is going to choose the "unwanted" action. This means that there would be no Nash equilibrium.

Moral Hazard (cont'd)

❖ Recall the principal's problem:

$$\min_{w_i \in \{w_1,\dots,w_n\}} \sum_i f_i(a) w_i$$

subject to

$$\sum_i f_i(a) u(w_i) - g(a) \geq \bar{u}$$

$$\sum_i f_i(a) u(w_i) - g(a) \geq \sum_i f_i(a_j) u(w_i) - g(a_j), \qquad \forall a_j \neq a$$

Using variable transformation, the problem becomes

$$\min_{u_i \in \{u_1,\dots,u_n\}} \sum_i f_i(a) v(u_i)$$

subject to

$$\sum_i f_i(a) u_i - g(a) \geq \bar{u}$$

$$\sum_i f_i(a) u_i - g(a) \geq \sum_i f_i(a_j) u_i - g(a_j), \qquad \forall a_j \neq a$$

Let λ and μ_j be the Lagrange multipliers on the two sets of constraints. The FOC with respect to u_i is

$$v'(u_i) = \lambda + \sum_j \mu_j \left(1 - \frac{f_i(a_j)}{f_i(a)}\right), \qquad \forall i$$

Suppose λ is the "base utility". Then,

$$f_i(a_j) < f_i(a) \;\; \Rightarrow \;\; \text{agent earns a "bonus" relative to his base utility}$$

Intuitively, if the principal's desired outcome is more likely to occur, then she should pay the agent more. By the same token,

$$f_i(a_j) > f_i(a) \;\; \Rightarrow \;\; \text{agent is "punished"}$$

- Note: the principal's problem can be thought of as a statistical inference problem. What the principal cares about is only how the outcome indicates what the agent has done. Since in reality,

  $$x_1 < \cdots < x_n \;\; \not\Rightarrow \;\; w_1 < \cdots < w_n$$

  we need the wage to be monotone, so that it's a good "estimate" of the agent's effort.

This gives a general result about the cost for the principal:

$$c^{SB}(a) > c^{FB}(a), \qquad \forall a \neq \arg\min_a g(a)$$

- Step 2: the principal solves

  $$\max_a \sum_i f_i(a) v(x_i) - c^{SB}(a) \;\; \rightarrow \;\; a^{SB}$$

- Example.
  - Two actions $a \in \{a_\ell, a_h\}$, $g(a_h) > g(a_\ell)$
  - Suppose the principal wants to implement a_h. Choose u_i such that
$$v'(u_i) = \lambda + \mu\left(1 - \frac{f_i(a_\ell)}{f_i(a_h)}\right)$$
  - When is w_i increasing i? $\rightarrow$ The likelihood ratios
$$\frac{f_1(a_\ell)}{f_1(a_h)}, \dots, \frac{f_n(a_\ell)}{f_n(a_h)}$$
must be monotone in i for w_i to be monotone in i. This is the ***monotone likelihood ratio property***.
$$w_i \uparrow \text{in } i \;\Leftarrow\; \frac{f_i(a_\ell)}{f_i(a_h)} \downarrow \text{in } i$$

- The characteristics of the optimal contract are
  - The agent is not fully insured
  - Basic trade-off: risk allocation v.s. incentive
  - Optimal wage scheme does not depend on the principal's benefit x_i
  - The agent's participation constraint is binding $\rightarrow$ the agent does not earn a rent. This implies that moral hazard does not induce labor market distortions. So the basic moral hazard model cannot explain the voluntary unemployment.

- Special case: agent is risk-neutral
  - If the agent is risk-neutral, then the principal can get the first-best.
  - Suppose the principal pays
$$w_i = x_i - P$$
$$u_A = \sum_i f_i(a)x_i - g(a) - P \geq \bar{u}$$
The agent is going to maximize total surplus.
    - Effectively, this is to say that the principal is selling the agent the technology, "sell the shop". So that the agent is the ***residual claimant*** for x_i.