Mechanism Design

CMPUT 654: Modelling Human Strategic Behaviour

S&LB §10.1-10.2

Logistics

- **Assignment #2** will be released on **Thursday**
- See the <u>course schedule</u> for paper presentation assignments
- Assignment #1 is about half-marked; should have results by the end of the week
- I will email solutions to Assignment #1 when it is marked; please **do not share the solutions** with anyone outside the class

Recap: Social Choice

Definition: A **social choice function** is a function $C : L^n \to O$, where

- $N = \{1,2,\ldots,n\}$ is a set of **agents**
- O is a finite set of **outcomes**
- L is the set of (non-strict) **total orderings** over O.

Definition: A **social welfare function** is a function $C : L^n \to L$, where N, O, and L are as above.

Notation:
We will denote **i's preference order** as $\succeq_i \in L$, and a **profile** of preference orders as $[\succeq] \in L^n$.

Recap:
Voting Scheme Properties

Definition:
W is **Pareto efficient** if for any $o_1, o_2 \in O$,

$$(\forall i \in N : o_1 \succ_i o_2) \implies (o_1 \succ_W o_2).$$

Definition:
W is **independent of irrelevant alternatives** if, for any $o_1, o_2 \in O$ and any two preference profiles $[\succ'], [\succ''] \in L$,

$$(\forall i \in N : o_1 \succ_i' o_2 \iff o_1 \succ_i'' o_2) \implies (o_1 \succ_{W[\succ']} o_2 \iff o_1 \succ_{W[\succ'']} o_2).$$

Definition:
W does not have a **dictator** if

$$\neg i \in N : \forall[\succ] \in L^n : \forall o_1, o_2 \in O : (o_1 \succ_i o_2) \implies (o_1 \succ_W o_2).$$

Recap: Arrow's Theorem

Theorem: (Arrow, 1951)
If $|O| > 2$, any social welfare function that is Pareto efficient and independent of irrelevant alternatives is dictatorial.

- Unfortunately, restricting to social choice functions instead of full social welfare functions doesn't help.

Theorem: (Muller-Satterthwaite, 1977)
If $|O| > 2$, any social choice function that is weakly Pareto efficient and monotonic is dictatorial.

Lecture Outline

1. Recap & Logistics
2. Mechanism Design with Unrestricted Preferences
3. Quasilinear Preferences
4. Paper scheduling

Mechanism Design

- In the social choice lecture, we assumed that agents report their preferences **truthfully**
- We now allow agents to report their preferences **strategically**
- Which social choice functions are **implementable** in this new setting?
    - **Question:** Wait, didn't we prove that social choice was hopeless?

Bayesian Game Setting

Definition:

A **Bayesian game setting** is a tuple (N, O, Θ, p, u) where

- N is a finite set of n **agents**,
- O is a set of **outcomes**,
- $\Theta = \Theta_1 \times \cdots \times \Theta_n$ is a set of possible **type profiles**,
- p is a **common prior** distribution over Θ, and
- $u = (u_1, \ldots, u_n)$, where $u_i : O \to \mathbb{R}$ is the **utility function** for player i.

This differs from a Bayesian game only in that utilities are defined on **outcomes** rather than **actions**, and agents are not (yet) endowed with an action set.

Mechanism

Definition:

A **mechanism** for a Bayesian game setting (N, O, Θ, p, u) is a pair (A, M), where

- $A = A_1 \times \cdots A_n$, where A_i is the set of **actions** available to agent i, and
- $M : A \to \Delta(O)$ maps each **action profile** to a distribution over **outcomes**

Intuitively, a **mechanism designer** (sometimes called **The Center**) needs to decide among outcomes in some Bayesian game setting, and so they design a mechanism that **implements** some social choice function.

Dominant Strategy Implementation

Definition:

Given a Bayesian game setting (N, O, Θ, p, u), a mechanism (A, M) is an **implementation in dominant strategies** of a social choice function C (over N and O) if,

1. The Bayesian game $(N, A, \Theta, p, u \circ M)$ induced by (A, M) has an equilibrium in dominant strategies, and

2. In any such equilibrium s^*, and for any type profile $\theta \in \Theta$, we have $M(s^*(\theta)) = C(u(\cdot, \theta))$.

Bayes-Nash Implementation

Definition:
Given a Bayesian game setting (N, O, Θ, p, u), a mechanism (A, M) is an **implementation in Bayes-Nash equilibrium** of a social choice function C (over N and O) if

1. There exists a Bayes-Nash equilibrium of the Bayesian game $(N, A, \Theta, p, u \circ M)$ induced by (A, M) such that
2. for every type profile $\theta \in \Theta$ and action profile $a \in A$ that can arise in equilibrium, $M(a) = C(u(\cdot, \theta))$.

The Space of All Mechanisms Is Enormous

- The space of all functions that map actions to outcomes is **impossibly large** to reason about
- **Question:** How could we ever prove that a given social choice function is **not implementable**?
- Fortunately, we can restrict ourselves without loss of generality to the class of **truthful, direct** mechanisms

Direct Mechanisms

Definition: A **direct** mechanism is one in which $A_i = \Theta_i$ for all agents $i \in N$.

Definition:
A direct mechanism is **truthful** (or **incentive compatible**) if, for all type profiles $\theta \in \Theta$, it is a dominant strategy in the game induced by the mechanism for each agent to report their true type.

Definition:
A direct mechanism is **Bayes-Nash incentive compatible** if there exists a Bayes-Nash equilibrium of the induced game in which every agent always truthfully reports their type.

Revelation Principle

Theorem: (Revelation Principle)

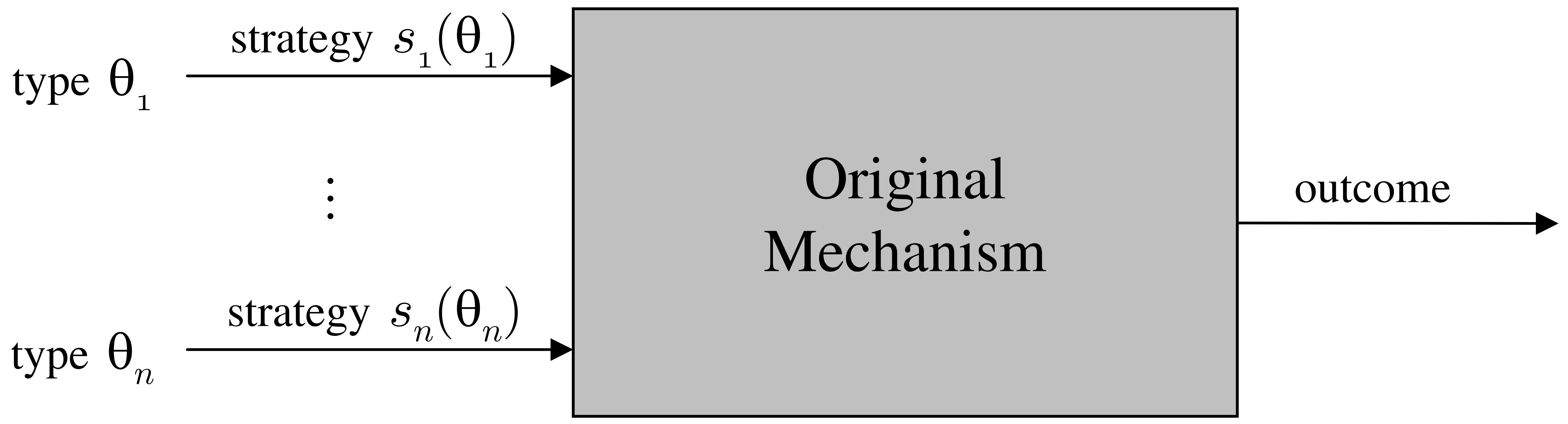
If there exists **any** mechanism that implements a social choice function C in dominant strategies, then there exists a **direct** mechanism that implements C in dominant strategies and is **truthful**.

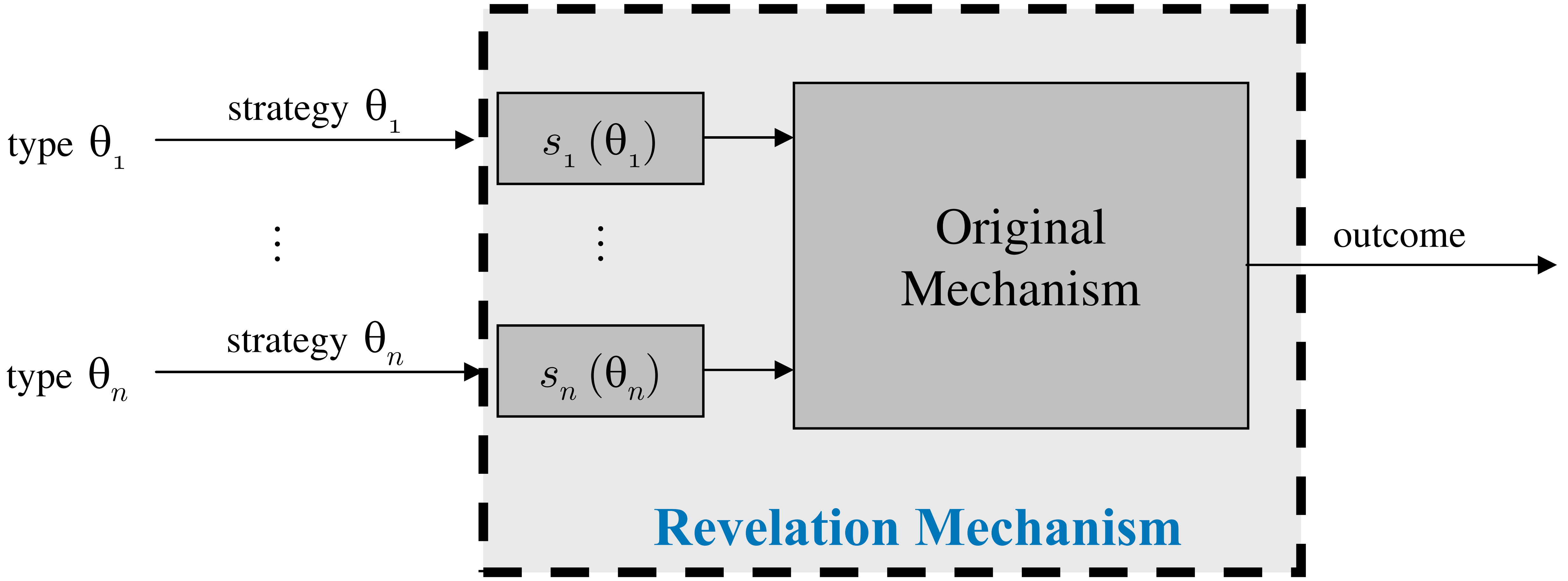
- Identical result for implementation in Bayes-Nash equilibrium

Revelation Principle Proof

1. Let (A, M) be an **arbitrary mechanism** that implements C in Bayesian game setting (N, O, Θ, p, u).
2. Construct the **revelation mechanism** $(\Theta, \overline{M})$ as follows:
   - For each type profile $\theta \in \Theta$, let $a^*(\theta)$ be the action profile in which every agent plays their dominant strategy in the game induced by (A, M).
   - Define $\overline{M}(\theta) = M(a^*(\theta))$.
3. Each agent reporting type $\hat{\theta}_i$ will yield the same outcome as every agent of type $\hat{\theta}_i$ playing their dominant strategy in M
4. So it is a dominant strategy for each agent to report their true type $\hat{\theta}_i = \theta_i$.

Revelation Mechanism





(Image: Shoham & Leyton-Brown 2008)

General
Dominant-Strategy Implementation

Theorem: (Gibbard-Satterthwaite)

Consider any social choice function C over N and O. If $|O| > 2$ (there are at least **three** outcomes),

1. C is **onto**; that is, for every outcome $o \in O$ there is a preference profile $[\succ]$ such that $C([\succ]) = o$ (this is sometimes called **citizen sovereignty**), and

2. C is dominant-strategy **truthful**,

then C is **dictatorial**.

Hold On A Second

Haven't we already seen an example of a dominant-strategy truthful direct mechanism?

Second Price Auction

- Outcomes are $O = \{(i \text{ gets object, pays } \$x) \mid i \in N, x \in \mathbb{R}\}$
- Types are $\theta_i = \mathbb{R}$, where an agent i with type $x \in \mathbb{R}$ has preferences:

$$(i \text{ gets object, pays } \$y') \succ_i (i \text{ gets object, pays } \$y'') \text{ for all } y' < y'' \text{ and } y' < x,$$
$$(i \text{ gets object, pays } \$y') \succ_i (j \text{ gets object, pays } \$y'') \quad \text{for all } y' < x \text{ and } i \neq j,$$
$$(j \text{ gets object, pays } \$y'') \succ_i (i \text{ gets object, pays } \$y') \quad \text{for all } y' > x \text{ and } i \neq j.$$

- Social choice function: Assign the item to the agent with the highest type
- Actions: Agents directly announce their type via sealed bid

- **Question:** Why is this not ruled out by Gibbard-Satterthwaite?

Restricted Preferences

- Gibbard-Satterthwaite only applies to social choice functions that operate on **every possible** preference ordering over the outcomes
- By **restricting the set of preferences** that we operate over, we can circumvent Gibbard-Satterthwaite

Quasilinear Preferences

Definition:

Agents have **quasilinear preferences** in an n-player Bayesian game setting when

1. the set of outcomes is $O = X \times \mathbb{R}^n$ for a finite set X,
2. the utility of agent i given type profile θ for an element $(x, p) \in O$ is $u_i\big((x, p), \theta\big) = v_i(x, \theta) - f_i(p_i)$, where
3. $v_i : X \times \Theta \to \mathbb{R}$ is an **arbitrary** function, and
4. $f_i : \mathbb{R} \to \mathbb{R}$ is a **monotonically increasing** function.

Quasilinear Preferences, informally

- **Intuitively:** Agents' preferences are split into
  1. finite set of **nonmonetary** outcomes (e.g., allocation of an object)
  2. monetary **payment** made to **The Center** (possibly negative)
- These two preferences are **linearly** related
- Agents are permitted **arbitrary preferences** over nonmonetary outcomes, but **not over payments**
- Agents care only about the **outcome selected** and their **own payment**
  - *and*, the amount they care about the outcome is **independent** of their payment

Direct Quasilinear Mechanism

Definition:

A **direct quasilinear mechanism** is a pair (χ, p), where

- $\chi : \Theta \to \Delta(X)$ is the **choice rule** (often called the **allocation rule**), which maps from a profile of reported types to a distribution over nonmonetary outcomes, and

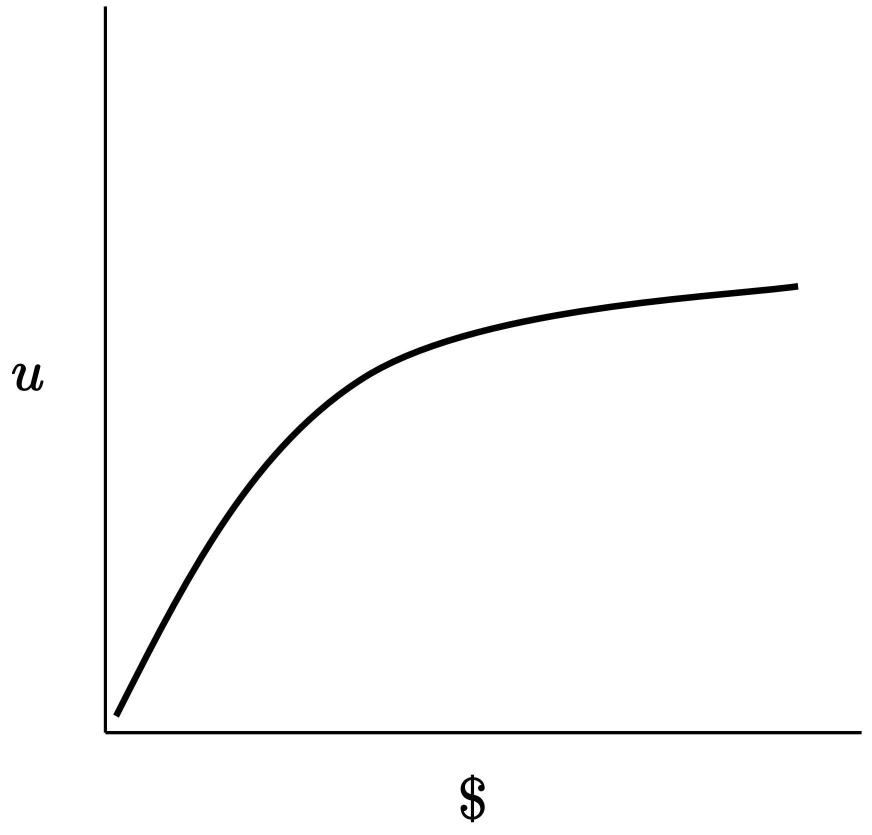
- $p : \Theta \to \mathbb{R}^n$ is the **payment rule**, which maps from a profile of reported types to a payment for each agent.

Value for Money

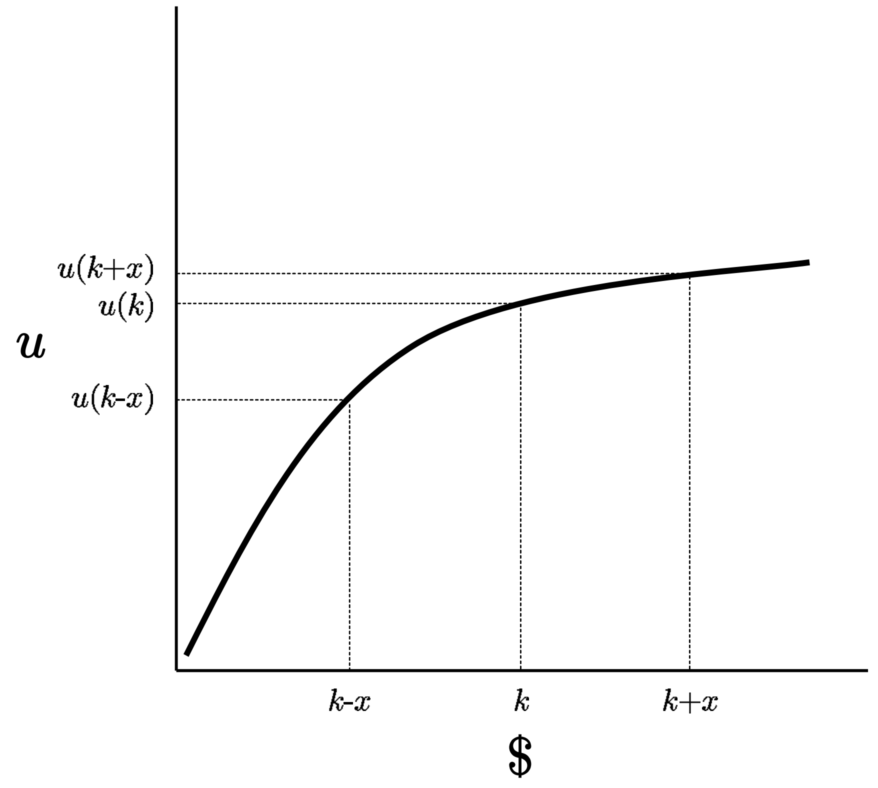
$$u_i\big((x,p),\theta\big) = v_i(x,\theta) - f_i(p_i)$$

- f_i represents agent i's **value for money**
  - **Question:** Why do we need a function instead of just a coefficient?
- The amount that you value \$1 will typically depend on how much money you **already have**:
  - An extra \$100 can change your life if you are starving
  - If you are a millionaire, you might not even notice the difference
- A **nonlinear** value for money can yield differing attitudes toward **risk**

Risk Aversion



(c) Risk aversion

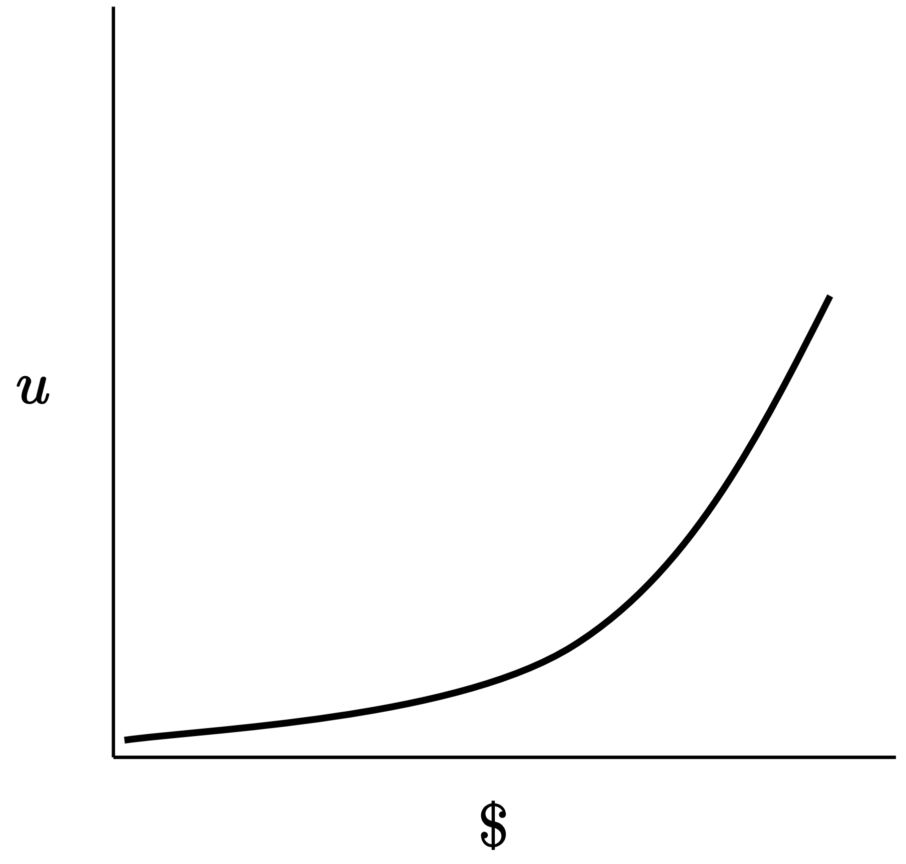


(d) Risk aversion: fair lottery

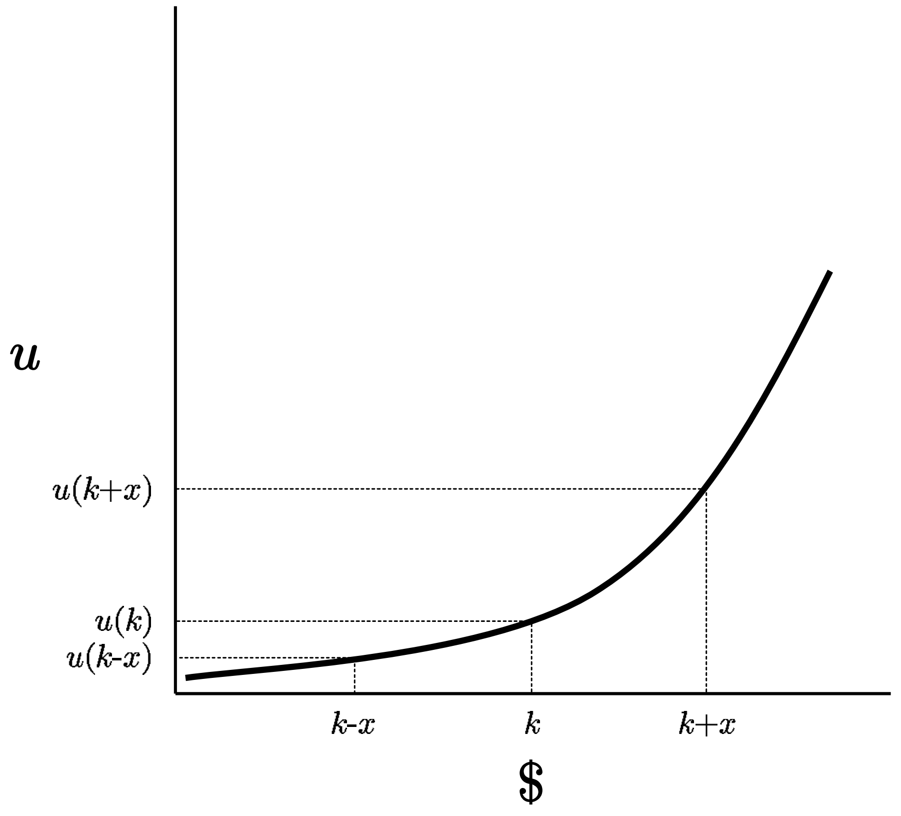
- A **concave** f_i models **decreasing marginal value** of money
- An agent with concave f_i is said to be **risk averse**, because they will **strictly prefer** to receive a lottery's **expected value** rather than to play the lottery
- **Question:** Is risk aversion irrational?

(Image: Shoham & Leyton-Brown 2008)

Risk Seeking



(e) Risk seeking

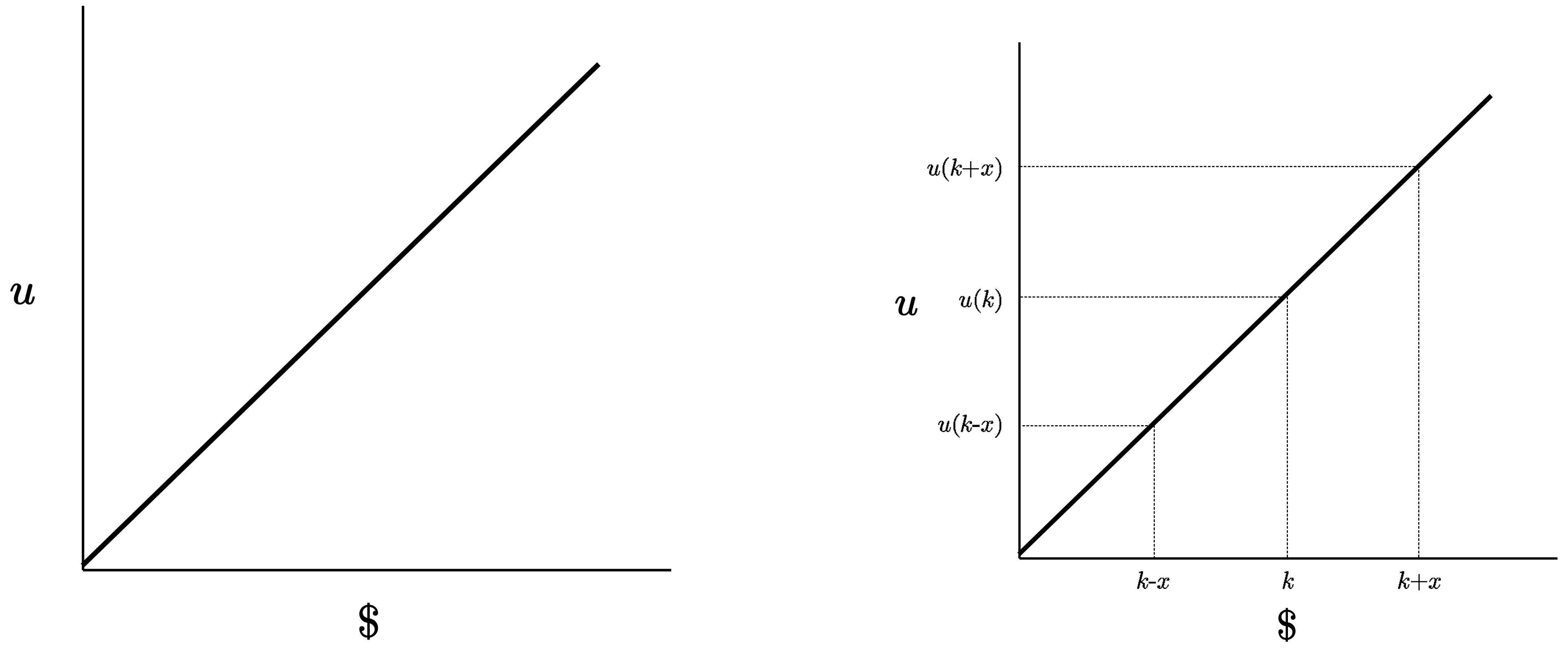


(f) Risk seeking: fair lottery

- A **convex** f_i models **increasing marginal value** of money
- An agent with convex f_i is said to be **risk seeking**, because they will **strictly prefer** to **play the lottery** rather than to receive a lottery's expected value
- **Question:** Is risk seeking irrational?

(Image: Shoham & Leyton-Brown 2008)

Risk Neutrality



(a) Risk neutrality

(b) Risk neutrality: fair lottery

- A **linear** f_i models **constant marginal value** of money
- An agent with linear f_i is said to be **risk neutral**, because they will be **indifferent** between receiving a lottery's **expected value** or playing the lottery

(Image: Shoham & Leyton-Brown 2008)

Transferable Utility

- Consider two agents i and j, who are both **risk-neutral**
- **Question:** Must they have the same value for money?
  - **No**, because they might have **different slopes:**

$$f_i(x) = \beta_i x$$
$$f_j(x) = \beta_j x$$
$$\beta_i \neq \beta_j$$

- When we additionally assume that $\beta_i = \beta_j$ for all $i, j \in N$, we say that the agents have **transferable utility**
  - Because I can increase i's utility by **exactly the amount** that I decrease j's utility, just by moving money from j to i
- Transferable utility is a **standard assumption** in quasilinear settings

Valuations

Definition:
A Bayesian game exhibits **conditional utility independence** if for all agents $i \in N$, all outcomes $o \in O$, and all pairs of joint types $\theta, \theta' \in \Theta$, it holds that $\theta_i = \theta'_i \implies u_i(o, \theta) = u_i(o, \theta')$.

- When this condition holds, we can write utility as $u_i(o, \theta_i)$
- Can equivalently refer to an agent's **valuation**: $v_i(x) = u_i(x, \theta_i)$.
- **Question:** When might this condition fail to hold?
- **Question:** Can we refer to an agent's valuation when this condition fails?

$$v_i(x) = u_i(x, \theta)$$

PAPER PRESENTATION SCHEDULING

Random dictatorship:

1. I have put the students into the random order on the right
2. We need to fill the timeslots in the spreadsheet
3. Every person chooses their favourite remaining slot, in order
4. You may steal an existing slot for a 2% penalty on your project
   - bumped person chooses immediately next
   - price for a paper increases by 2% every time it is stolen

Questions:

1. Is random dictatorship **dominant strategy truthful**?
2. Is the **full procedure with stealing** DS truthful?
3. Is this procedure social welfare maximizing?

Summary

- **Mechanism design**: Setting up a system for **strategic agents** to provide input to a **social choice function**
- **Revelation Principle** means we can restrict ourselves to **truthful direct** mechanisms without loss of generality
- Non-dictatorial dominant-strategy mechanism design is **impossible in general** (Gibbard-Satterthwaite)
- The special case of **quasi-linear preferences** will allow us to **circumvent** Gibbard-Satterthwaite (next time!)