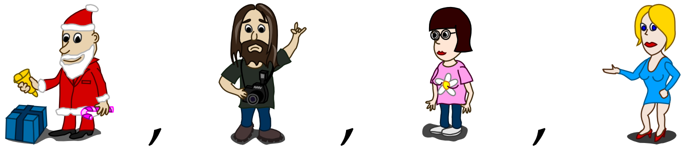
Implementation Theory

Can we design a game, that yields a particular set of outcomes as equilibria?

Implementation Theory

Outcomes $C = \{$  $\}$

Preference Profile $\succsim = (\succsim_i) = (\succsim_1, \succsim_2, \cdots, \succsim_n) \in \mathcal{P}$

Choice Rule $f(\succsim)$: To return some candidates:

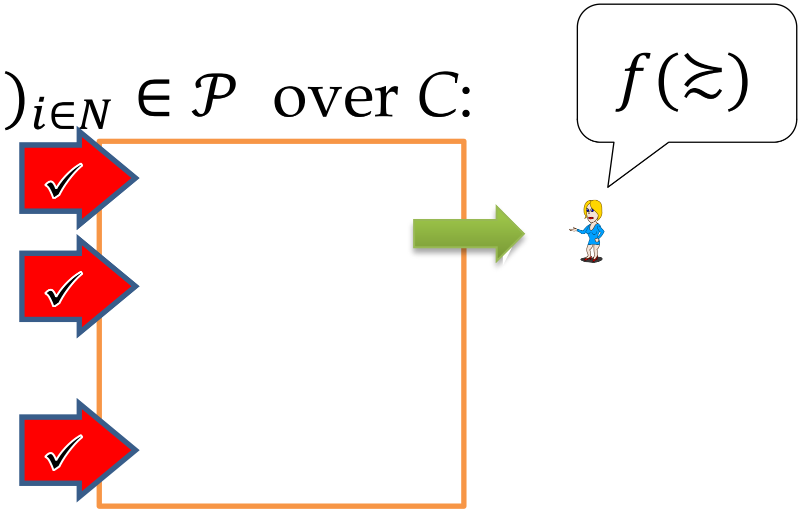
Voters' Preferences $\succsim = (\succsim_i)_{i \in N} \in \mathcal{P}$ over C:

Voter 1: $\succsim_1 \quad \succsim_1 \quad \succsim_1$

Voter 2: $\succsim_2 \quad \succsim_2 \quad \succsim_2$

…

Voter n: $\succsim_n \quad \succsim_n \quad \succsim_n$



The Implementation Problem

- The set N of n players
- The set C of feasible outcomes
- The set $\mathcal{P}$ of all preference profiles over C

- **Choice function** $f: \mathcal{P} \to C$

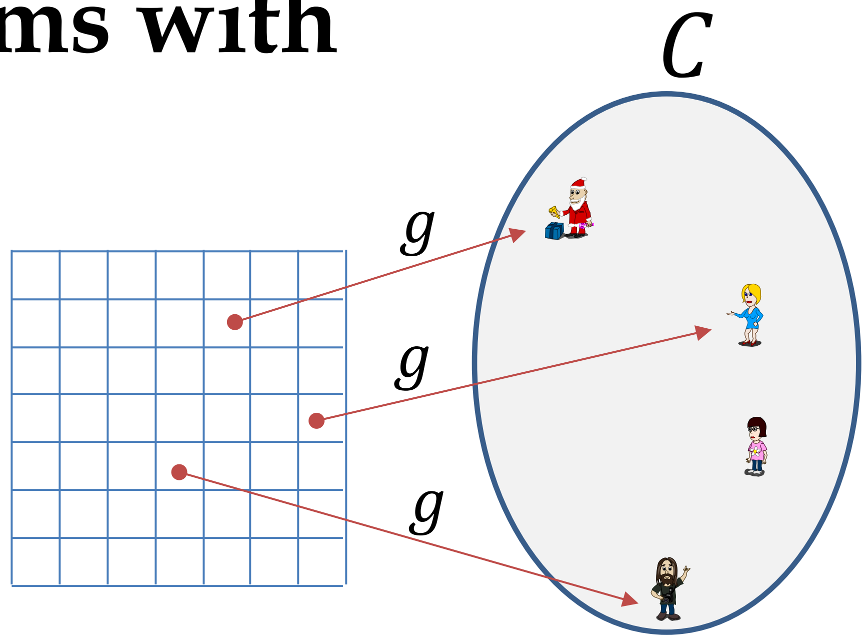
  $f(\succsim^1) = f(\succsim_1^1, \dots, \succsim_n^1) = a \in C.$

  $f(\succsim^2) = f(\succsim_1^2, \dots, \succsim_n^2) = b \in C.$

  And so on …

Choice Rule:
Generally, many outcomes are returned.

Strategic Game Forms with Consequences in C



Actions for player i:

$A_i = \{\succsim_i^1, \succsim_i^2, \succsim_i^3, \dots\}$

A **strategic game form with consequences in** C is a triple

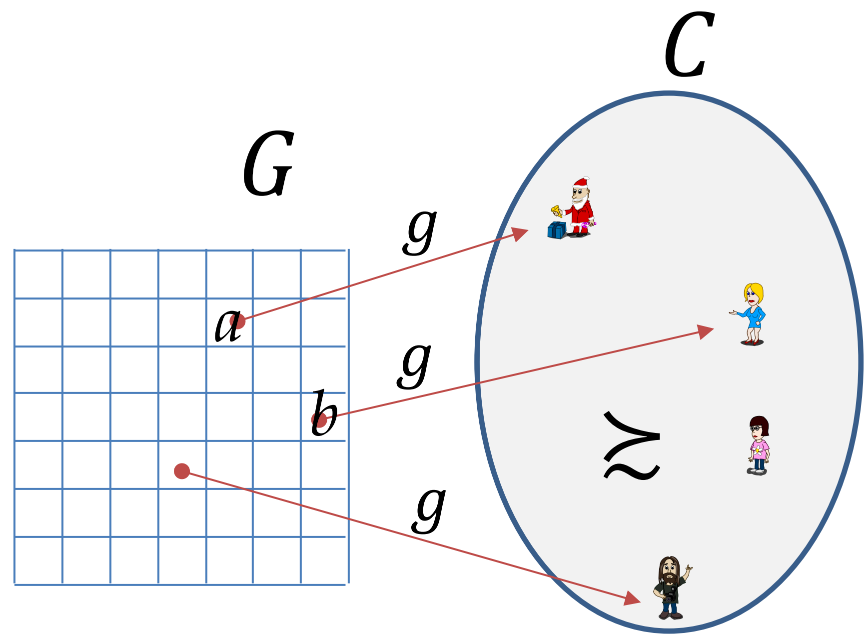
$$G = \langle N, (A_i), g \rangle$$

where $g: A \to C$ is an **outcome function**.

Players N

Actions $A_i = \{\succsim_i^1, \succsim_i^2, \succsim_i^3, \dots\}$

Outcome function $g: A \to C$



$$G = \langle N, (A_i), g \rangle \in \mathcal{G} \qquad \text{and} \qquad \succsim = (\succsim_i)_{i \in N} \in \mathcal{P}$$

G and $\succsim$ induce a strategic game $\langle N, (A_i), (\succsim_i') \rangle$

$$\langle N, (A_i), (\succsim_i') \rangle = \langle G, \succsim \rangle \in \mathcal{G} \times \mathcal{P}$$

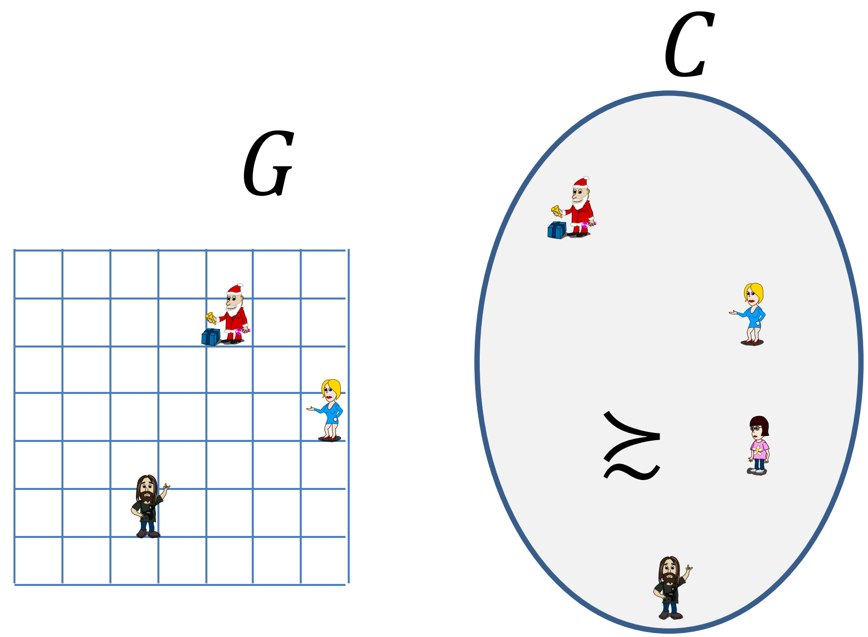
where for each $i \in N$, $\succsim_i'$ is defined by

$a \succsim_i' b$ if and only if $g(a) \succsim_i g(b)$.

Players N

Actions $A_i = \{\succsim_i^1, \succsim_i^2, \succsim_i^3, \ldots\}$

Outcome function $g: A \to C$



$G = \langle N, (A_i), g \rangle \in \mathcal{G}$ and $\succsim = (\succsim_i)_{i \in N} \in \mathcal{P}$

G and $\succsim$ induce a strategic game $\langle N, (A_i), (\succsim_i') \rangle$

$\langle N, (A_i), (\succsim_i') \rangle = \langle G, \succsim \rangle \in \mathcal{G} \times \mathcal{P}$

where for each $i \in N$, $\succsim_i'$ is defined by

$a \succsim_i' b$ if and only if $g(a) \succsim_i g(b)$.

Extensive Game Forms with Consequences in C

It is possible to implement choice rules using extensive game form with consequences in C.

In this course we concentrate our discussion on the implementation of choice rules using strategic game form with consequences in C.

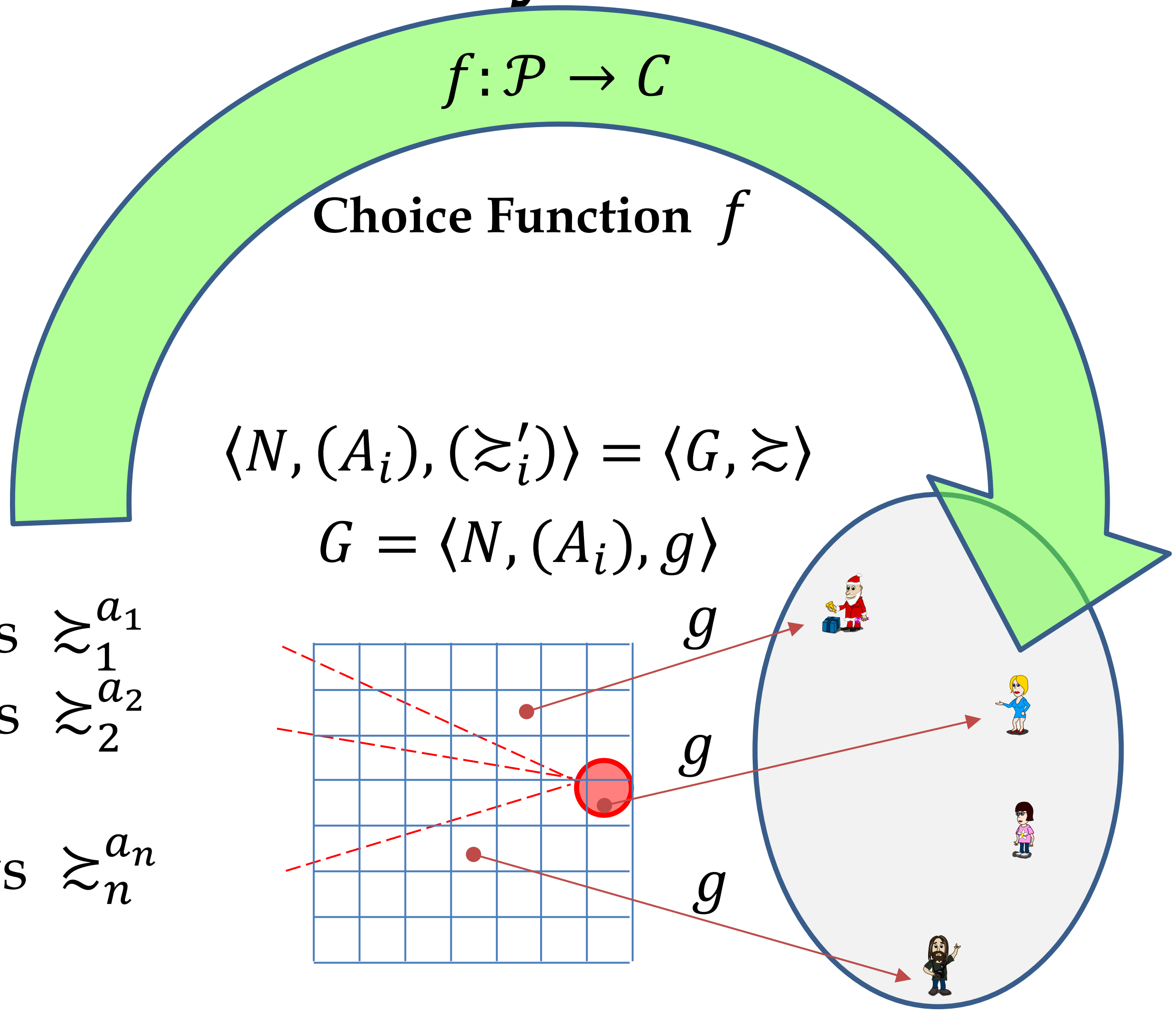
Environments

The planner operates in an **environment** $\langle N, C, \mathcal{P}, \mathcal{G} \rangle$:

- a finite set N of players, with $|N| \geq 2$
- a set C of outcomes
- a set $\mathcal{P}$ of preference profiles over C
- a set $\mathcal{G}$ of game forms with consequences in C

Planner's task: to design a game form $G = \langle N, (A_i), g \rangle$ such that the outcomes of $\langle G, \succsim \rangle \in \mathcal{G} \times \mathcal{P}$, for each $\succsim = (\succsim_i)_{i \in N} \in \mathcal{P}$, coincide with $f(\succsim)$.

Implementation Theory



$$\langle N, (A_i), (\succsim_i') \rangle = \langle G, \succsim \rangle$$

$$G = \langle N, (A_i), g \rangle$$

Voter 1: $\succsim_1$, plays $\succsim_1^{a_1}$

Voter 2: $\succsim_2$, plays $\succsim_2^{a_2}$

…

Voter n: $\succsim_n$, plays $\succsim_n^{a_n}$

Solution Concepts

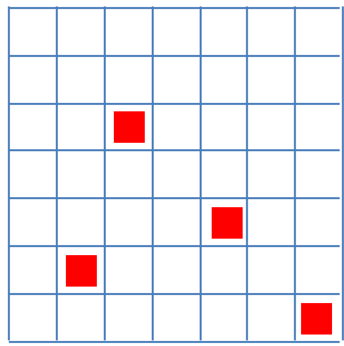
For the environment $\langle N, C, \mathcal{P}, \mathcal{G} \rangle$

$$\left.\begin{array}{l}\text{Game form } G \in \mathcal{G} \\ \text{Preference Profile } \succsim \in \mathcal{P}\end{array}\right\} \langle G, \succsim \rangle \in \mathcal{G} \times \mathcal{P}$$

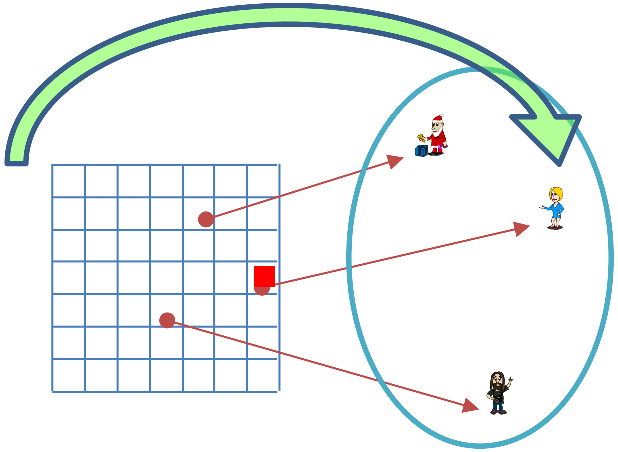
A **solution concept** $\mathcal{S}(G, \succsim) \subseteq A$ of G and $\succsim$ is a set-valued function $\mathcal{S}: \mathcal{G} \times \mathcal{P} \to 2^A$.

Examples of 'Solution Concept' $\mathcal{S} \subseteq A$

- *Nash equilibrium outcomes*
- *Social welfare maximising outcomes*
- *And so on…*



Solution Concepts



EXAMPLE. $\mathcal{S} = \mathrm{NE}$

$\mathcal{S}(G, \succsim) = \mathrm{NE}(G, \succsim)$ is the set of all NEs in game $\langle G, \succsim \rangle = \langle N, (A_i), (\succsim_i') \rangle$.

If $g(\mathrm{NE}(G, \succsim)) = f(\succsim)$ for any $\succsim$, we say that game form $G = \langle N, (A_i), g \rangle$ with outcome function g **NE-implements** the choice rule f; f is hence **NE-implementable** in $\langle N, C, \mathcal{P}, \mathcal{G} \rangle$.

$\mathcal{S}$-Implementation and $\mathcal{S}$-Implementability of Choice Rules

DEFINITION. Let $\langle N, C, \mathcal{P}, \mathcal{G}\rangle$ be an environment and let $\mathcal{S}$ be a solution concept. The game form $G = \langle N, (A_i), g\rangle \in \mathcal{G}$ with outcome function g is said to **$\mathcal{S}$-implement** the choice rule $f: \mathcal{P} \to C$ if for every preference profile $\succsim \in \mathcal{P}$ we have

$$g\big(\mathcal{S}(G, \succsim)\big) = f(\succsim).$$

In this case we say the choice rule f is **$\mathcal{S}$-implementable** in $\langle N, C, \mathcal{P}, \mathcal{G}\rangle$.

$\mathcal{S}$-Implementability

The key point is to *design* a $G \in \mathcal{G}$ with a g that $\mathcal{S}$-implements the choice rule $f: \mathcal{P} \to C$.

Given $\succsim \in \mathcal{P}$.

- We specify the intended winner(s) by $f(\succsim)$.
- **Example**. Can we find a game form $G = \langle N, (A_i), g \rangle \in \mathcal{G}$, for which the Nash equilibrium outcomes (i.e., $\mathcal{S} = \mathrm{NE}$) correspond (i.e., $g\big(\mathcal{S}(G, \succsim)\big)$) to the intended winner(s) $f(\succsim)$?

There is another notion of implementation theory: the action of a player is a preference profile $\succsim \in \mathcal{P}$. That is, each player must announce a preference relation for every player.

Environment $\langle N, C, \mathcal{P}, \mathcal{G} \rangle$

Actions $A_1 = A_2 = \cdots = A_n = \mathcal{P}$

Consider game form $G = \langle N, (A_i), g \rangle \in \mathcal{G}$.

If for any $\succsim \in \mathcal{P}$, and an outcome $a^* = (\succsim, \dots, \succsim)$

- $a^* = (\succsim, \dots, \succsim) \in \mathcal{S}(G, \succsim)$, and
- $g(a^*) = g((\succsim, \dots, \succsim)) \in f(\succsim)$.

Then we say

G **truthfully** $\mathcal{S}$**-implements** the choice rule f;

f is truthfully $\mathcal{S}$**-implementable** in $\langle N, C, \mathcal{P}, \mathcal{G} \rangle$.

Truthful $\mathcal{S}$-Implementation

DEFINITION. Let $\langle N, C, \mathcal{P}, \mathcal{G} \rangle$ be an environment in which $\mathcal{G}$ is a set of strategic game forms for which the set of actions of each player i is a set $\mathcal{P}$ of preference profiles, and let $\mathcal{S}$ be a solution concept. The game form $G = \langle N, (A_i), g \rangle \in \mathcal{G}$ **truthfully $\mathcal{S}$-implements** the choice rule $f : \mathcal{P} \to C$ if for every preference profile $\succsim \in \mathcal{P}$ we have

(cont.)

- $a^* \in \mathcal{S}(G, \succsim)$ where $a_i^* = \succsim$ for each $i \in N$ (every player reporting the true preference profile is a solution of the game)

- $g(a^*) \in f(\succsim)$(the outcome if every player reports the true preference profile is a member of $f(\succsim)$).

G **truthfully** $\mathcal{S}$**-implements** the choice rule f.
f is **truthfully** $\mathcal{S}$**-implementable** in environment $\langle N, C, \mathcal{P}, \mathcal{G} \rangle$.

Truthful $\mathcal{S}$-Implementation

Note 1. This notion requires the set of actions of each player to be the set of preference profiles and 'truth telling' to be a solution to every game.

- *The set of actions of each player* i *is a set* $\mathcal{P}$ *of preference profiles.*
- $a^* \in \mathcal{S}(G, \succsim)$ *where* $a_i^* = (\succsim, \ldots, \succsim)$.

Truthful $\mathcal{S}$-Implementation

Note 2. It allows (non truth-telling) solutions to yield outcomes that are inconsistent with the choice rule.

- $a^* \in \mathcal{S}(G, \succsim)$ *where* $a_i^* = \succsim$ *for each* $i \in N$.
- $g(a^*) \in f(\succsim)$
- *Therefore, it is possible that there is an action profile* $b \in \mathcal{S}(G, \succsim)$ *where* $b_i \neq \succsim$ *for some* $i \in N$, *as long as* $g(b) \notin f(\succsim)$.

Truthful $\mathcal{S}$-Implementation

Note 3. It allows there to be preference profiles for which not every outcome prescribed by the choice rule corresponds to a solution.

- $a^* \in \mathcal{S}(G, \succsim)$ *where* $a_i^* = \succsim$ *for each* $i \in N$.
- $g(a^*) \in f(\succsim)$
- *Therefore, we can have a preference profile* $\succsim$ *and* $s \in f(\succsim)$ *for which there is no* $a \in \mathcal{S}(G, \succsim)$ *such that* $g(a) = s$.

Implementation in Dominant Strategies

We want to design a game, so that the outcomes $f(\gtrsim)$ that we wish to implement are consistent with the solution concept of *dominant strategy equilibrium* (DSE).

That is, $\mathcal{S} = \mathrm{DSE}$. We design a game, such that $g(\mathrm{DSE}(G, \gtrsim)) = f(\gtrsim)$.

Dominant Strategy Equilibria of Strategic Games

DEFINITION. A **dominant strategy equilibrium of a strategic game** $\langle N, (A_i), (\succsim_i) \rangle$ is a profile $a^* \in A$ of actions with the property that for every player $i \in N$ we have $(a_{-i}, a_i^*) \succsim_i a$ for all $a \in A$.

Here a_i^* is known as a dominant strategy of $i \in N$.
Q: what is the intuitive meaning of a dominant strategy?

Is it possible to find a game form $G = \langle N, (A_i), g \rangle$ that DSE-implements a choice rule $f: \mathcal{P} \to C$?

That is, is it possible to find a game form $G = \langle N, (A_i), g \rangle$ such that for every preference profile $\succsim \in \mathcal{P}$, we have $g(\text{DSE}(G, \succsim)) = f(\succsim)$?

To answer this question, we need a new concept about the choice rule f.

Dictatorial Choice Rules

Consider a special choice rule f. Under this choice rule f, there is a (lucky) player $j \in N$, such that no matter what other players declare to prefer, the outcome is always what player j considers to be the best.

Dictatorial Choice Rules

Formally, if there is a (lucky) player $j \in N$ such that for any preference profile $\succsim \in \mathcal{P}$ and outcome $a \in f(\succsim)$ we have $a \succsim_j b$ for all $b \in C$.

The choice rule f is said to be **dictatorial** and the player j is called a **dictator**.

Gibbard-Satterthwaite Theorem

PROPOSITION. (*Gibbard-Satterthwaite Theorem*) Let $\langle N, C, \mathcal{P}, \mathcal{G} \rangle$ be an environment in which C contains at least three members, $\mathcal{P}$ is the set of all possible preference profiles, and $\mathcal{G}$ is the set of strategic game forms. Let $f: \mathcal{P} \to C$ be a choice rule that is DSE-implementable and satisfies the condition

for every $a \in C$ *there exists* $\succsim \in \mathcal{P}$ *such that* $f(\succsim) = \{a\}$.

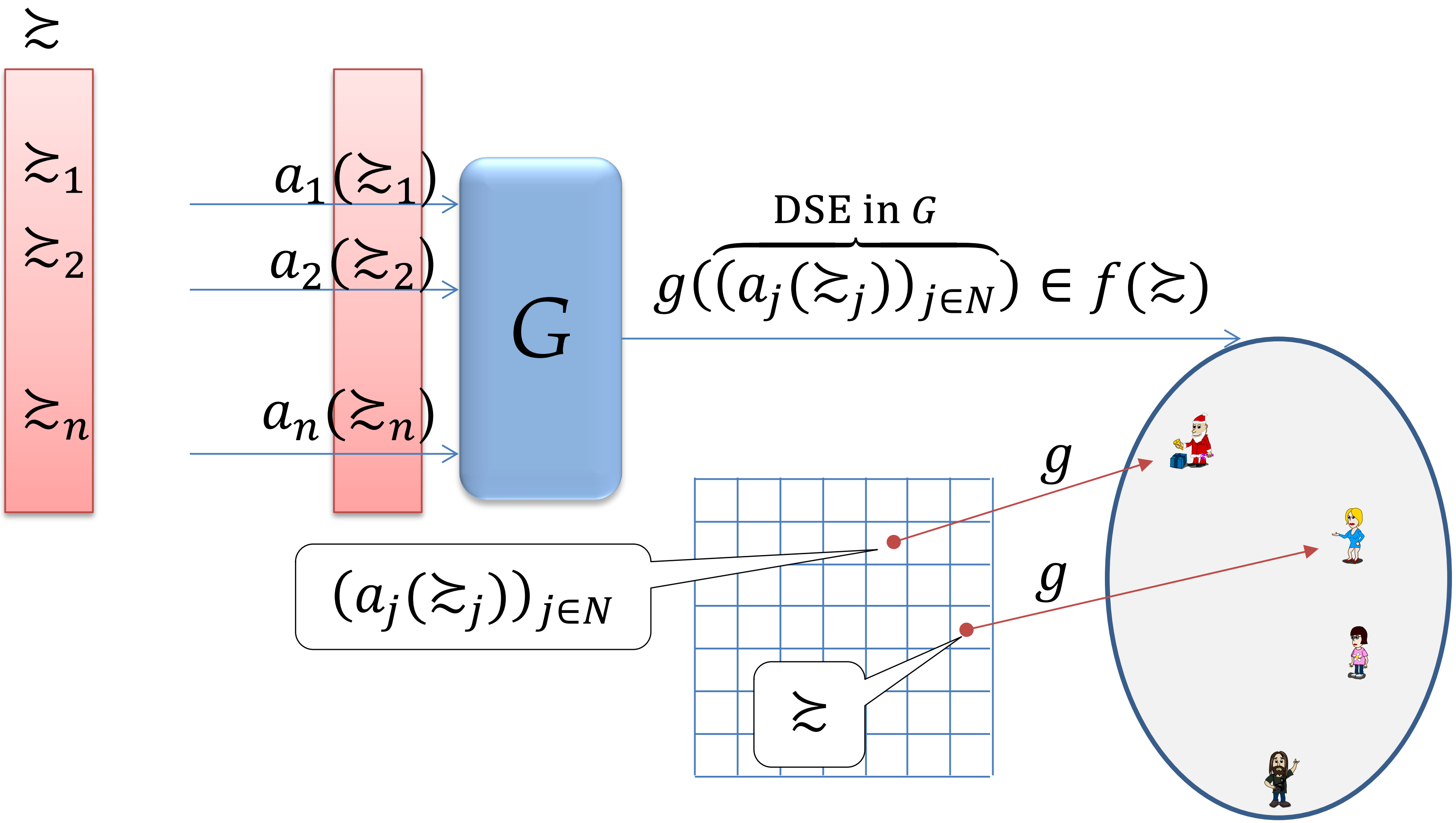
Then f is dictatorial.

We now continue to introduce another important concept, known as the **Revelation Principle for DSE-implementation**.

Consider that $G \in \mathcal{G}$ DSE-implements $f: \mathcal{P} \to C$

Player i's preference is $\succsim_i$. Let $a_i(\succsim_i)$ denote a dominant action for $i \in N$ in $\langle G, \succsim \rangle$. Player i will play $a_i(\succsim_i)$.

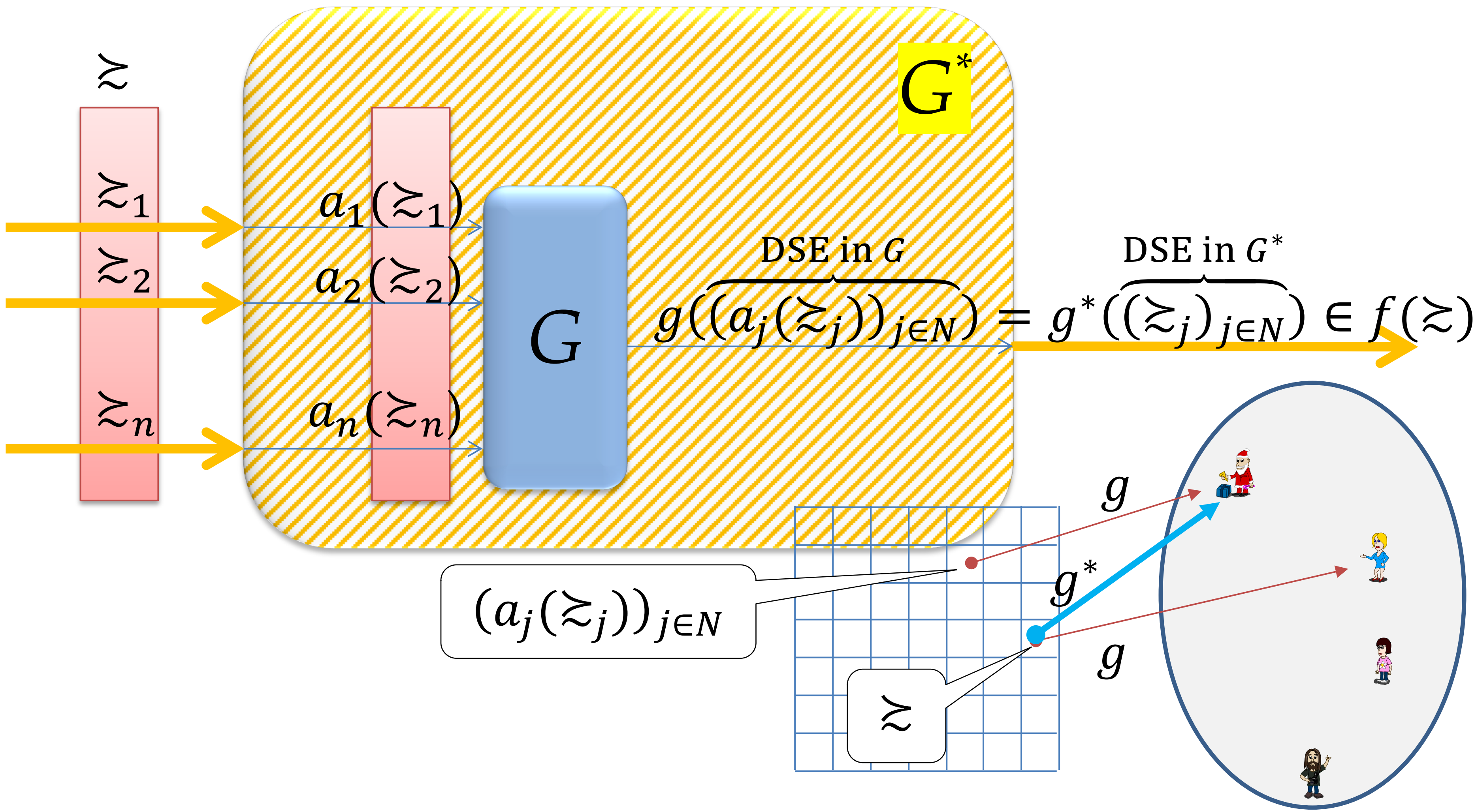
Hence, $(a_i(\succsim_i))_{i \in N} = (a_1(\succsim_1), a_2(\succsim_2), \dots, a_n(\succsim_n))$ is a DSE, and $g((a_i(\succsim_i))_{i \in N}) \in f(\succsim)$. But player i needs to pay efforts to find what $a_j(\succsim_j)$ is.



$\succsim$
$\succsim_1$
$\succsim_2$
$\succsim_n$
$a_1(\succsim_1)$
$a_2(\succsim_2)$
$a_n(\succsim_n)$
G
DSE in G
$g(\overbrace{(a_j(\succsim_j))_{j\in N}}^{\text{DSE in } G}) \in f(\succsim)$
g
g
$(a_j(\succsim_j))_{j\in N}$
$\succsim$

G DSE-implements f by $g((a_i(\succsim_i))_{i\in N})$

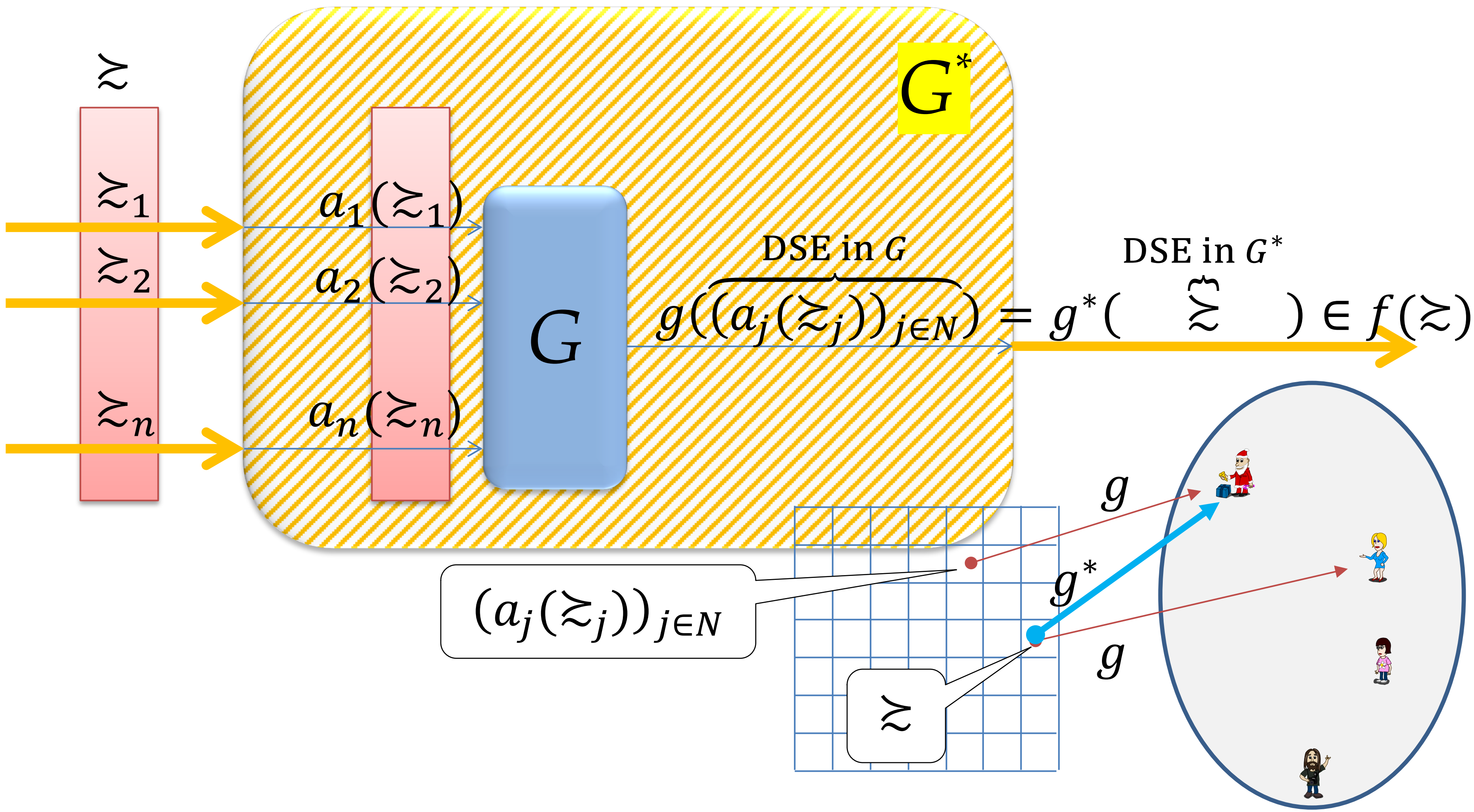
We define a <u>new</u> game form $G^* = \langle N, (A_i), g^* \rangle \in \mathcal{G}$. G^* is the same as G, except that g is replaced by g^*, and $g^*(\succsim) = g((a_i(\succsim_i))_{i\in N})$.


$\succsim$
G^*
$\succsim_1$
$\succsim_2$
$\succsim_n$
$a_1(\succsim_1)$
$a_2(\succsim_2)$
$a_n(\succsim_n)$
G
DSE in G
DSE in G^*
$g\big(\overbrace{(a_j(\succsim_j))_{j\in N}}^{\text{DSE in } G}\big) = g^*\big(\overbrace{(\succsim_j)_{j\in N}}^{\text{DSE in } G^*}\big) \in f(\succsim)$
g
g^*
g
$(a_j(\succsim_j))_{j\in N}$
$\succsim$

G DSE-implements f by $g((a_i(\succsim_i))_{i \in N})$

$$G^* = \langle N, (A_i), g^* \rangle \in \mathcal{G}.$$
$$g^*(\succsim) = g((a_i(\succsim_i))_{i \in N}).$$

If every player i plays $\succsim_i$ in $\langle G^*, \succsim \rangle$, the outcome is the same as every player i plays $a_i(\succsim_i)$ in $\langle G, \succsim \rangle$. Note that $\succsim_i$ is readily known to every player i, but player i needs to find what $a_j(\succsim_j)$ is.

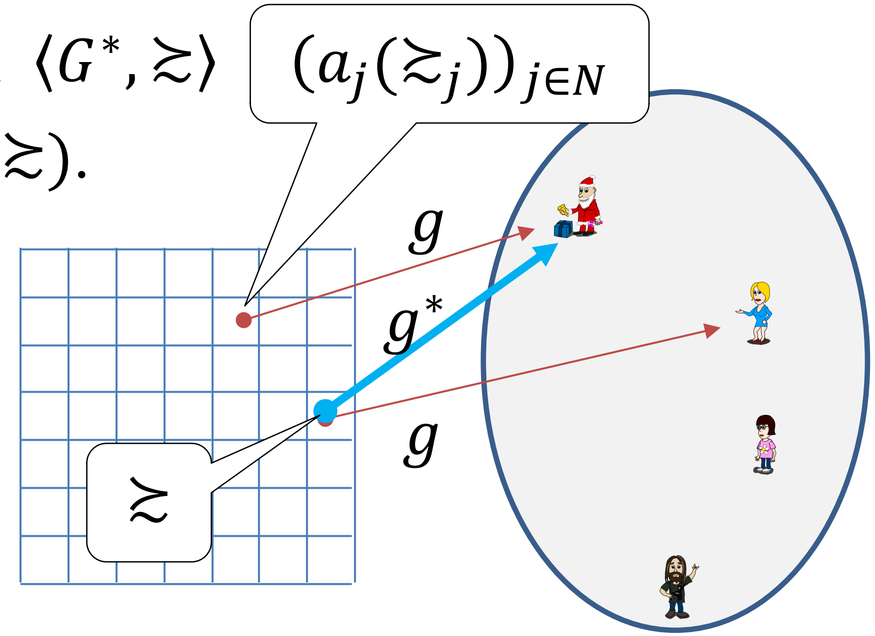

$\succsim$
$\succsim_1$
$\succsim_2$
$\succsim_n$
G^*
$a_1(\succsim_1)$
$a_2(\succsim_2)$
$a_n(\succsim_n)$
G
DSE in G
DSE in G^*
$g(\overbrace{(a_j(\succsim_j))_{j\in N}}^{\text{DSE in } G}) = g^*(\overbrace{\succsim}^{\text{DSE in } G^*}) \in f(\succsim)$
g
g^*
g
$(a_j(\succsim_j))_{j\in N}$
$\succsim$

G DSE-implements f by $g((a_i(\succsim_i))_{i \in N})$

If every player i plays $\succsim_i$ in $\langle G^*, \succsim \rangle$, the outcome is the same as every player i plays $a_i(\succsim_i)$ in $\langle G, \succsim \rangle$.

Hence, $\succsim$ must be a DSE in $\langle G^*, \succsim \rangle$ and $g^*((\succsim_i)_{i \in N}) = g^*(\succsim) \in f(\succsim)$.

Everyone just plays $\succsim_i$ in $\langle G^*, \succsim \rangle$. That will be a DSE.



G DSE-implements f by $g((a_i(\succsim_i))_{i\in N})$

G^* DSE-implements f by $g^*((\succsim_i)_{i\in N})$

Now we consider $\langle G', \succsim \rangle$ in which the set of actions of each player is $\mathcal{P}$ ($A_i = \mathcal{P}$, $A = \times_{i\in N} \mathcal{P}$), and the outcome function g' is given by

$$g'((\overbrace{\succsim(1)}^{\text{Player } 1'\text{s action}}, \overbrace{\succsim(2)}^{\text{Player } 2'\text{s action}}, \dots, \overbrace{\succsim(n)}^{\text{Player } n'\text{s action}})) \\ = g^*((\succsim_1(1), \succsim_2(2), \dots, \succsim_n(n)))$$

where $\succsim(i) \in \mathcal{P}$ is an action (a preference profile) played by player $i \in N$.

Therefore, it will be a DSE for everyone to play $\succsim$.

$$
\begin{aligned}
&g'((\overbrace{\succsim}^{\text{Player } 1'\text{s action}}, \overbrace{\succsim}^{\text{Player } 2'\text{s action}}, \dots, \overbrace{\succsim}^{\text{Player } n'\text{s action}})) \\
&= g^*((\succsim_1, \succsim_2, \dots, \succsim_n)) \\
&= g((a_i(\succsim_i))_{i \in N})
\end{aligned}
$$

G' truthfully DSE-implements f by $g'(\succsim, \dots, \succsim)$.

Revelation Principle for DSE-implementation

LEMMA. Let $\langle N, C, \mathcal{P}, \mathcal{G} \rangle$ be an environment in which $\mathcal{G}$ is the set of strategic game forms. If a choice rule $f : \mathcal{P} \to C$ is DSE-implementable then

a. f is truthfully DSE-implementable

b. there is a strategic game form $G^* = \langle N, (A_i), g^* \rangle \in \mathcal{G}$ in which A_i is the set of all preference *relations* (rather than profiles) such that for all $\succsim \in \mathcal{P}$, the action profile $\succsim$ is a DSE of the strategic game $\langle G^*, \succsim \rangle$ and $g^*(\succsim) \in f(\succsim)$.

If a choice rule f cannot be truthfully DSE-implemented then it cannot be DSE-implemented.

Discussion

Is the following choice rule DSE-implementable?

$$f(\succsim) = \begin{cases} a & \text{if for all } i \in N \text{ we have } a \succ_i b \text{ for all } b \neq a \\ a^* & \text{otherwise.} \end{cases}$$

a^* is an arbitrary outcome in C.

Discussion

$$f(\succsim) = \begin{cases} a & \text{if for all } i \in N \text{ we have } a \succ_i b \text{ for all } b \neq a \\ a^* & \text{otherwise.} \end{cases}$$

Consider the situation

$$x \succ_1 a \succ_1 a^* \text{for all } x \notin \{a, a^*\}$$

$$a \succ_i x \text{ for all } i \neq 1 \text{for all } x$$

Hence $f(\succsim) = a^*$ but $f(\succsim_{-1}, \succsim_1') = a$.

$$a \succ_1' x \succ_1' a^* \text{for all } x \notin \{a, a^*\}$$

Therefore, $\succsim_1$ is not a dominant strategy for player 1 in $\langle G^*, \succsim \rangle$, as it is worse than $\succsim_1'$.

Gibbard-Satterthwaite Theorem

PROPOSITION. (*Gibbard-Satterthwaite Theorem*) Let $\langle N, C, \mathcal{P}, \mathcal{G} \rangle$ be an environment in which C contains at least three members, $\mathcal{P}$ is the set of *<u>all</u>* possible preference profiles, and $\mathcal{G}$ is the set of strategic game forms. Let $f: \mathcal{P} \to C$ be a choice rule that is DSE-implementable and satisfies the condition

$$\text{for every } a \in C \text{ there exists } \succsim \in \mathcal{P}$$
$$\text{such that } f(\succsim) = \{a\}$$

Then f is dictatorial.

Given

C contains at least three members,

$\mathcal{P}$ is the set of <u>*all*</u> possible preference profiles,

and $\mathcal{G}$ is the set of strategic game forms.

If we want to have

- $f: \mathcal{P} \to C$ is DSE-implementable;
- for every $a \in C$, $\exists \succsim \in \mathcal{P}$ s.t. $f(\succsim) = \{a\}$;
- f is not dictatorial.

This is simply **impossible**!

Shall We Build a Swimming Pool?

Possible outcomes in C are all in the form of

$$(x, (m_i)_{i \in N}),\ x = 0{,}1.$$

Player i's utility: $u_i = \theta_i x - m_i$ for some $\theta_i \in \mathbb{R}$.

(Each $(u_i)_{i \in N}$ represents a preference profile—hence $\mathcal{P}$ does not contain *all* possible preference profiles.)

Choice rule to implement: $f: \mathbb{R}^n \to C$

such that $x = 1$ if and only if $\sum_{i \in N} \theta_i \geq \gamma \geq 0$.

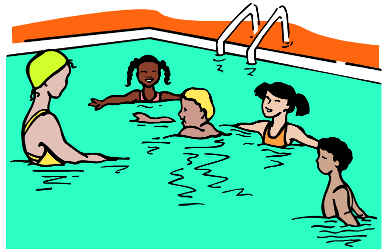
Groves Mechanism

Groves Mechanism is the game form $\langle N, (A_i), g \rangle$.
(Also known as the **Clarke-Groves game form**.)

Each player $j \in N$ declare his (possibly fake) value $\hat{\theta}_j \in A_j = \mathbb{R}$. For $\hat{\theta} = (\hat{\theta}_i) \in A$, $x(\hat{\theta}) = 1$ if and only if $\sum_{i \in N} \hat{\theta}_i \geq \gamma \geq 0$, and $g(\hat{\theta}) = (x(\hat{\theta}), m(\hat{\theta}))$, where

$$m_j(\hat{\theta}) = \begin{cases} h_j\left(\hat{\theta}_{-j}\right) & x(\hat{\theta}) = 0 \\ (\gamma - \sum_{i \in N \setminus \{j\}} \hat{\theta}_i) + h_j(\hat{\theta}_{-j}) & x(\hat{\theta}) = 1 \end{cases}$$

EXAMPLE.



$N = \{1,2\}$, $\theta_1 = 1.5$, $\theta_2 = 0.8$, $\gamma = 2$.

$$\hat{\theta}_1 = 1.5, \hat{\theta}_2 = 0.8$$

	θ_i	$\hat{\theta}_i$	m_i	u_i
Player 1	1.5	1.5	$1.2 + h_1(\hat{\theta}_{-1})$	$0.3 - h_1(\hat{\theta}_{-1})$
Player 2	0.8	0.8	$0.5 + h_2(\hat{\theta}_{-2})$	$0.3 - h_2(\hat{\theta}_{-2})$
Sum	2.3	2.3		

Therefore, $x(\hat{\theta}) = 1$. If a player i lies ($\hat{\theta}_i \neq \theta_i$), as long as $x(\hat{\theta}) = 1$, $m_i(\hat{\theta}) = m_i(\theta)$ is unchanged. The outcome is unchanged.

EXAMPLE.

$N = \{1,2\},\ \theta_1 = 1.5,\ \theta_2 = 0.8,\ \gamma = 2.$

$\hat{\theta}_1 = 1.5, \hat{\theta}_2 = 0.2$

	θ_i	$\hat{\theta}_i$	m_i	u_i
Player 1	1.5	1.5	$h_1(\hat{\theta}_{-1})$	$-h_1(\hat{\theta}_{-1})$
Player 2	0.8	0.2	$h_2(\hat{\theta}_{-2})$	$-h_2(\hat{\theta}_{-2})$
Sum	2.3	1.7		

Therefore, $x(\hat{\theta}) = 0$. Both players suffer. There is no point 'lying too little.'

EXAMPLE.

$N = \{1,2\},\ \theta_1 = 0.5,\ \theta_2 = 0.8,\ \gamma = 2.$

$\hat{\theta}_1 = 1.3,\ \hat{\theta}_2 = 0.8$

	θ_i	$\hat{\theta}_i$	m_i	u_i
Player 1	0.5	1.3	$1.2 + h_1(\hat{\theta}_{-1})$	$-0.7 - h_1(\hat{\theta}_{-1})$
Player 2	0.8	0.8	$0.7 + h_2(\hat{\theta}_{-2})$	$0.1 - h_2(\hat{\theta}_{-2})$
Sum	1.3	2.1		

Therefore, $x(\hat{\theta}) = 1$. Actually there is no point for player 1 to tell the lie.

Why It Is Not a Good Idea Not to Tell the Truth Under the Groves Mechanism

$$m_j(\hat{\theta}) = \begin{cases} h_j(\hat{\theta}_{-j}) & x(\hat{\theta}) = 0 \\ (\gamma - \sum_{i\in N\setminus\{j\}} \hat{\theta}_i) + h_j(\hat{\theta}_{-j}) & x(\hat{\theta}) = 1 \end{cases}$$

Would it not be better for player *j* to tell a lie $\hat{\boldsymbol{\theta}}_j \neq \boldsymbol{\theta}_j$?

Case 1.

If $x(\theta_{-j}, \theta_j) = x(\theta_{-j}, \hat{\theta}_j)$, then $m_j(\theta_{-j}, \theta_j) = m_j(\theta_{-j}, \hat{\theta}_j)$, and hence $g(\theta_{-j}, \theta_j) = g(\theta_{-j}, \hat{\theta}_j)$.

Therefore, it is not better for player *j* to tell a lie $\hat{\theta}_j \neq \theta_j$.

$$m_j(\hat{\theta}) = \begin{cases} h_j(\hat{\theta}_{-j}) & x(\hat{\theta}) = 0 \\ (\gamma - \sum_{i\in N\setminus\{j\}} \hat{\theta}_i) + h_j(\hat{\theta}_{-j}) & x(\hat{\theta}) = 1 \end{cases}$$

$$u_j(\hat{\theta}) = \theta_j - m_j(\hat{\theta})$$

Case 2. $\sum_{i\in N} \theta_i < \gamma$, hence $x(\theta_{-j}, \theta_j) = 0$.

$x(\theta_{-j}, \theta_j) = 0 \quad ➔ \quad u_j(\theta_{-j}, \theta_j) = -h_j(\theta_{-j})$

$x(\theta_{-j}, \hat{\theta}_j) = 1 \quad ➔$

$u_j(\theta_{-j}, \hat{\theta}_j) = \theta_j - (\gamma - \sum_{i\in N\setminus\{j\}} \theta_i + h_j(\theta_{-j}))$

$u_j(\theta_{-j}, \hat{\theta}_j) = (-\gamma + \sum_{i\in N} \theta_i) - h_j(\theta_{-j}))$

$u_j(\theta_{-j}, \hat{\theta}_j) < -h_j(\theta_{-j})$

Therefore, it is not better for player j to tell a lie $\hat{\theta}_j \neq \theta_j$.

$$m_j(\hat{\theta}) = \begin{cases} h_j(\hat{\theta}_{-j}) & x(\hat{\theta}) = 0 \\ (\gamma - \sum_{i \in N \setminus \{j\}} \hat{\theta}_i) + h_j(\hat{\theta}_{-j}) & x(\hat{\theta}) = 1 \end{cases}$$

$$u_j(\hat{\theta}) = \theta_j - m_j(\hat{\theta})$$

Case 3. $\sum_{i \in N} \theta_i \geq \gamma$, hence $x(\theta_{-j}, \theta_j) = 1$.

$x(\theta_{-j}, \theta_j) = 1$ ➔

$$u_j(\theta_{-j}, \theta_j) = \theta_j - (\gamma - \sum_{i \in N \setminus \{j\}} \theta_i + h_j(\theta_{-j}))$$

$$u_j(\theta_{-j}, \theta_j) = (-\gamma + \sum_{i \in N} \theta_i) - h_j(\theta_{-j}))$$

$$u_j(\theta_{-j}, \theta_j) \geq -h_j(\theta_{-j})$$

$x(\theta_{-j}, \hat{\theta}_j) = 0$ ➔ $u_j(\theta_{-j}, \hat{\theta}_j) = -h_j(\theta_{-j})$

Therefore, it is not better for player j to tell a lie $\hat{\theta}_j \neq \theta_j$.

$$m_j(\hat{\theta}) = \begin{cases} h_j\left(\hat{\theta}_{-j}\right) & x(\hat{\theta}) = 0 \\ (\gamma - \sum_{i \in N \setminus \{j\}} \hat{\theta}_i) + h_j(\hat{\theta}_{-j}) & x(\hat{\theta}) = 1 \end{cases}$$

$$u_j\left(\hat{\theta}\right) = \theta_j - m_j(\hat{\theta})$$

Case 1. $x(\theta_{-j}, \theta_j) = x(\theta_{-j}, \hat{\theta}_j)$.

Case 2. $x(\theta_{-j}, \theta_j) = 0,\ x(\theta_{-j}, \hat{\theta}_j) = 1$.

Case 3. $x(\theta_{-j}, \theta_j) = 1,\ x(\theta_{-j}, \hat{\theta}_j) = 0$.

In all cases, it is not better for player *j* to tell a lie $\hat{\theta}_j \neq \theta_j$.

Hence a dominant action for each player j is to choose $\hat{\theta}_j = \theta_j$. Then $g(\theta) = f(\theta)$, so that $\langle N, (A_i), g \rangle$ truthfully DSE-implements *f*.

Groves Mechanism

PROPOSITION. Let $\langle N, C, \mathcal{P}, \mathcal{G} \rangle$ be an environment in which $C = \{(x, m) : x \in \{0,1\}$ and $m \in \mathbb{R}^n\}$, $\mathcal{P}$ is the set of profiles $(\succsim_i)_{i \in N}$ in which each $\succsim_i$ is represented by a utility function of the form $\theta_i x - m_i$ for some $\theta_i \in \mathbb{R}$, and $\mathcal{G}$ is the set of strategic game forms; identify $\mathcal{P}$ with $\mathbb{R}^n$. A choice function $f : \mathbb{R}^n \to C$ with $f(\theta) = (x(\theta), m(\theta))$ for which

- $x(\theta) = 1$ if and only if $\sum_{i \in N} \theta_i \geq \gamma$
- for each $j \in N$ there is a function h_j such that $m_j(\theta) = x(\theta)(\gamma - \sum_{i \in N \setminus \{j\}} \theta_i) + h_j(\theta_{-j})$ for all $\theta \in \mathbb{R}^n$.

is truthfully DSE-implemented by the Groves mechanism $\langle N, (A_i), g \rangle$.

Clarke Tax

Clarke Tax is a special instance of Groves Mechanism.

Now suppose a group of agents vote whether to build a swimming pool or not.

A solution is to make an agent pay *tax* if its vote changes the outcome: its tax is related to how much its vote lowers the others' utility. Agents that do not end up changing the outcome do not pay any tax.

The Clarke Tax Algorithm

- Every agent $i \in N$ reveals its valuation $\hat{\theta}_j(x)$ for every possible outcome x.
- The social choice is $x^* = \arg\max_{\mathrm{x}} \sum_{i \in N} \hat{\theta}_i(x)$. Every agent *j* is levied a tax:

$$\mathrm{tax}_j = \sum_{i \in N \setminus \{j\}} \hat{\theta}_i(x^*) - \sum_{i \in N \setminus \{j\}} \hat{\theta}_i \left(\arg\max_{\mathrm{x}} \sum_{k \in N \setminus \{j\}} \hat{\theta}_k(x) \right)$$

Question: is tax_j positive or negative?

a_j	True worth of each outcome			Sum for each state without a_j			Tax for a_j
	x_1	x_2	x_3	x_1	x_2	x_3	
a_1	27	–33	6	–46	*23	*23	0
a_2	–36	12	24	*17	–22	5	–12
a_3	–9	24	–15	–10	–34	*44	0
a_4	–18	–15	33	–1	*5	–4	–9
a_5	17	2	–19	–36	–12	*48	0
Sum	–19	–10	*29				

Is this a Groves Mechanism?

Can a_1 and a_5 beneficially collude by untruthfully reveal their utilities?