Behavioral Implementation under Incomplete Information

Mehmet Barlo[1] Nuh Aygün Dalkıran[2]

May 1, 2019

---

[1] Sabancı University
[2] Bilkent University

Main Question

What shall a **planner** do if she/he were to **implement** a goal when the relevant **information** is **distributed** among "**predictably irrational**" individuals?

What we do?

We investigate **the implementation problem** under **incomplete information** when individuals' choices **need not be rational**.

Our results

- are complementary to "**Behavioral Implementation**" **de Clippel (2014)** [*AER*, 104(10): 2975-3002], which investigates the same problem under **complete information**.
- provide an important leap as **information asymmetries** are **inescapable** in many economic settings.

Motivation for Behavioral Implementation

People have cognitive limitations and are prone to **behavioral biases**:

- the attraction effect
- the status-quo bias
- intransitive indifference
- limited attention
- the endowment effect
- temptation and self-control
- compromise and framing effects ...

- Therefore, individual behavior may **not** be consistent with a **rational** preference relation.
- **Behavioral economics** and **psychology** offers **insights** on the **systematic** deviations from rationality.

Motivation for Behavioral Implementation

There is a recent trend towards **behavioral implementation**, i.e., **taking into account behavioral insights**.

- The Behavioral Insights Team (BIT) (a.k.a the **Nudge Unit**) was established in 2010 in the U.K.
  - "in order to improve the government policies using ideas drawn from the behavioral sciences"
- Thaler and Sunstein (2008) "**Nudge**: Improving Decisions about Health, Wealth, and Happiness "
- Australia, Canada, Germany, India, Indonesia, Ireland, Jordan, Netherlands, Peru, Singapore, Turkey, and the US among others started applying behavioral insights to their policies and programs.
- International institutions such as the EU, OECD, UN, World Bank established behavioral insights units to support their programs.

Literature moving in this direction

"We should take behavioral insights into account when designing mechanisms:"

- Borgers and Li (2018): Strategically Simple Mechanisms
- Li (2017): Obviously Strategy Proof Mechanisms
- De Clippel (2014): **Behavioral Implementation** (complete information),
- Korpela (2012): Implementation without Rationality Assumptions, (complete information)
- Spiegler (2011): Bounded Rationality and Industrial Organization,
- Bernheim (2009), Bernheim and Rangel (2009): Foundations for Behavioral Welfare Economics,
- Eliaz (2002): Fault Tolerant Implementation, (complete information)
- Hurwicz (1986): On the Implementation ... in Irrational Societies, (complete information)
- among many many others...

Example of a behavioral bias: Attraction Effect

Attraction Effect: Decoy alternatives –alternatives that are known to be dominated by another alternative– can cause preference reversals when they are introduced in the consideration set.

- The Economist
  - Option **A**: Online subscription: \$59;
  - Option **B**: Print subscription: \$125; →**decoy**
  - Option **C**: Online & Print subscription: \$125.

- **A** is expected to be chosen from {**A**, **C**} whereas
- **C** is expected to be chosen from the set {**A**, **B**, **C**}.

Attraction Effect in a policy-making context

Herne (1997) demonstrates how the presence of a decoy alternative causes attraction effect in a **policy-making context**:

- In 1993, Finland took the decision of building a new nuclear power plant to parliamentary vote.
- The majority of the **opponents** of nuclear power favored decentralized solar power plants as an alternative.
- The **proponents** of the nuclear power plant used coal power plants as a point of comparison to nuclear.

Attraction Effect in a policy-making context

Two dimensions seems relevant: **environment** and **reliability**.

- Nuclear power **dominates** coal as it is more environment friendly and reliable (stability/price)
- Solar power is **better** for the environment when compared to both nuclear and coal.
- High costs of solar panels and intermittency makes it **less** appealing against both nuclear and coal (reliability).

Coal is **dominated by nuclear** in **both dimensions** but it is **dominated by solar** power only in **environment dimension**.

(asymmetric domination)

Choices displaying Attraction Effect

It is expected by the proponents that

- *nuclear* is chosen from the grand set $\{coal, nuclear, solar\}$
- *solar* is chosen from the set $\{nuclear, solar\}$.

This violates weak axiom of revealed preferences (WARP)
[Sen (1971): WARP = Sen's α + Sen's β.]

- Sen's α (**independence of irrelevant alternatives**) fails!
- Sen's β (**expansion consistency**) fails as well!

Motivating Example: Alice and Bob

Alice and Bob are to decide what type of energy to employ or jointly invest in, be it coal energy, nuclear energy, or solar energy.

- Their individual choices **may not be rational**, i.e., they may violate weak axiom of revealed preferences.

Let the set of all relevant states regarding individual choices be Θ.

- There is incomplete information between Alice and Bob regarding the true state of the world $\theta \in \Theta$.

Motivating Example: Choices

- $X = \{coal, nuclear, solar, \}$,
- $\Theta = \Theta_A \times \Theta_B$ where $\Theta_i = \{\rho_i, \gamma_i\}$ for both $i \in \{A, B\}$.
- That is, $\Theta = \{(\rho_A, \rho_B), (\rho_A, \gamma_B), (\gamma_A, \rho_B), (\gamma_A, \gamma_B)\}$
- If the true state of the world is $\theta = (\theta_A, \theta_B)$
  - Alice only learns $\theta_A \in \Theta_A$.
  - Bob only learns $\theta_B \in \Theta_B$.

Motivating Example: Choices

Suppose that the individual choice behavior of Alice and Bob at each state of the world are **predicted** to be as follows:

	$C_A^{(\rho_A,\rho_B)}$	$C_B^{(\rho_A,\rho_B)}$	$C_A^{(\rho_A,\gamma_B)}$	$C_B^{(\rho_A,\gamma_B)}$	$C_A^{(\gamma_A,\rho_B)}$	$C_B^{(\gamma_A,\rho_B)}$	$C_A^{(\gamma_A,\gamma_B)}$	$C_B^{(\gamma_A,\gamma_B)}$
$\{c,n,s\}$	$\{n\}$	$\{s\}$	$\{n\}$	$\{n\}$	$\{n\}$	$\{c\}$	$\{c,s\}$	$\{n,s\}$
$\{c,n\}$	$\{n\}$	$\{n\}$	$\{n\}$	$\{n\}$	$\{n\}$	$\{c\}$	$\{n\}$	$\{c\}$
$\{c,s\}$	$\{c,s\}$	$\{s\}$	$\{s\}$	$\{s\}$	$\{c\}$	$\{c\}$	$\{c\}$	$\{s\}$
$\{n,s\}$	$\{n\}$	$\{s\}$	$\{s\}$	$\{s\}$	$\{n,s\}$	$\{n,s\}$	$\{s\}$	$\{s\}$

Possible Preferences and Behavioral Biases

	$C_A^{(\rho_A,\rho_B)}$	$C_B^{(\rho_A,\rho_B)}$	$C_A^{(\rho_A,\gamma_B)}$	$C_B^{(\rho_A,\gamma_B)}$	$C_A^{(\gamma_A,\rho_B)}$	$C_B^{(\gamma_A,\rho_B)}$	$C_A^{(\gamma_A,\gamma_B)}$	$C_B^{(\gamma_A,\gamma_B)}$
$\{c, n, s\}$	$\{n\}$	$\{s\}$	$\{n\}$	$\{n\}$	$\{n\}$	$\{c\}$	$\{c, s\}$	$\{n, s\}$
$\{c, n\}$	$\{n\}$	$\{n\}$	$\{n\}$	$\{n\}$	$\{n\}$	$\{c\}$	$\{n\}$	$\{c\}$
$\{c, s\}$	$\{c, s\}$	$\{s\}$	$\{s\}$	$\{s\}$	$\{c\}$	$\{c\}$	$\{c\}$	$\{s\}$
$\{n, s\}$	$\{n\}$	$\{s\}$	$\{s\}$	$\{s\}$	$\{n, s\}$	$\{n, s\}$	$\{s\}$	$\{s\}$
	$n \succ_A c \sim_A s$	$s \succ_B n \succ_B c$	**Att “c”**	**Att “c”**	**S-quo “c”**	$c \succ_B s \sim_B n$	**Cycles**	**?**

Motivating Example: The Goal

Suppose a social planner cares about the **welfare** of these agents in the **behavioral welfare economics** sense à la Bernheim and Rangel (2009):

- "An alternative **x** is strictly unambiguously chosen over another alternative **z** iff **z** is never chosen whenever **x** is available."
  - **x** does **not** have to be chosen but it **prevents** **z** to be chosen.
- "An alternative **x** is weakly unambiguously chosen over another alternative **z** iff whenever they are both available, **z** is never chosen unless **x** is chosen as well."
  - if **z** is chosen, then **x** is chosen as well —whenever they are both available.

Motivating Example: Generalized P.O.

Generalized Pareto Optimality à la Bernheim and Rangel (2009): An alternative x is strictly generalized Pareto-optimal if there does not exists y such that

- y is weakly unambiguously chosen over x for every agent and
- y is strictly unambiguously chosen over x for some agent.

Below, we refer to **strictly generalized Pareto-optimal alternatives** as **BR-optimal** alternatives.

Motivating Example: BR-Optima

Therefore, such a social planner, **without knowing the true state of the world**, would like the agents to collectively end up according to the following Social Choice Rule:

State:	(ρ_A, ρ_B)	(ρ_A, γ_B)	(γ_A, ρ_B)	(γ_A, ρ_B)
SCR: BR-Optimal	$\{n \quad , \quad s\}$	$\{n \quad , \quad s\}$	$\{c \quad , \quad n\}$	$\{c \quad , \quad s\}$

	$C_A^{(\rho_A,\rho_B)}$	$C_B^{(\rho_A,\rho_B)}$	$C_A^{(\rho_A,\gamma_B)}$	$C_B^{(\rho_A,\gamma_B)}$	$C_A^{(\gamma_A,\rho_B)}$	$C_B^{(\gamma_A,\rho_B)}$	$C_A^{(\gamma_A,\gamma_B)}$	$C_B^{(\gamma_A,\gamma_B)}$
$\{c, n, s\}$	$\{n\}$	$\{s\}$	$\{n\}$	$\{n\}$	$\{n\}$	$\{c\}$	$\{c, s\}$	$\{n, s\}$
$\{c, n\}$	$\{n\}$	$\{n\}$	$\{n\}$	$\{n\}$	$\{n\}$	$\{c\}$	$\{n\}$	$\{c\}$
$\{c, s\}$	$\{c, s\}$	$\{s\}$	$\{s\}$	$\{s\}$	$\{c\}$	$\{c\}$	$\{c\}$	$\{s\}$
$\{n, s\}$	$\{n\}$	$\{s\}$	$\{s\}$	$\{s\}$	$\{n, s\}$	$\{n, s\}$	$\{s\}$	$\{s\}$

Motivating Example: BR-Optima

It is customary to work with **Social Choice Sets** as social choice rules under incomplete information, see e.g., Jackson (1991), Bergemann and Morris (2008) among many others.

- A Social Choice Set is simply a **set** of social choice functions
- Each social choice function can be considered as a **state contingent allocation**.

We will work with the following "mutually exhaustive" social choice set:

State:	(ρ_A, ρ_B)	(ρ_A, γ_B)	(γ_A, ρ_B)	(γ_A, ρ_B)
SCR: BR-Optimal	$\{n \quad , \quad s\}$	$\{n \quad , \quad s\}$	$\{c \quad , \quad n\}$	$\{c \quad , \quad s\}$
$\mathbf{f}$	n	n	n	s
$\mathbf{f}'$	s	s	c	c

Social Choice Set: $F = \{\mathbf{f}, \mathbf{f}'\}$

Motivating Example: A mechanism

Let's try the following mechanism:

$$\mu = (\mathcal{M}, g)$$

$$\mathcal{M}_A = \{U, M, D\} \text{ and } \mathcal{M}_B = \{L, M, R\}$$

Bob

	g	L	M	R
Alice	U	n	c	n
	M	c	s	c
	D	n	s	s

Motivating Example: Opportunity Set

- Sugden (2004): The **opportunity set** of an agent is the set of alternatives from which he is unrestricted to choose.
- In case of a mechanism, a message sent by an individual **restricts** the **alternatives** the other agents can **generate**.
- De Clippel (2014): The **opportunity set** of agent i given m_{-i} under the mechanism $\mu = (\mathcal{M}, g)$ is given by

$$O_i^{\mu}(m_{-i}) := \{g(m_i, m_{-i}) | m_i \in \mathcal{M}_i\} \subseteq X.$$

Motivating Example: Equilibrium

De Clippel (2014): "An **equilibrium outcome** is among the chosen options within the set of outcomes he can generate through **unilateral** deviations."

Bob

		L	M	R
	U	n	c	n
Alice	M	c	s	c
	D	n	s	s

Ex: $O^{\mu}_{A}(\mathbf{L}) = \{\mathbf{n}, \mathbf{c}\}$; $O^{\mu}_{B}(\mathbf{D}) = \{\mathbf{n}, \mathbf{s}\}$.

m^* is a **Nash equilibrium of μ at θ** if $g(m^*(\theta)) \in C_i^{\theta}(O_i^{\mu}(m^*_{-i}))$ $\forall i$.

Choice Replies at (γ_A, γ_B)

Bob

Alice

$$\begin{array}{c|ccc} & L & M & R \\ \hline U & n & c & n \\ M & c & s & c \\ D & n & s & s \end{array}$$

Choices	$C_A^{(\gamma_A,\gamma_B)}$	$C_B^{(\gamma_A,\gamma_B)}$
$\{c, n, s\}$	$\{c, s\}$	$\{n, s\}$
$\{c, n\}$	$\{n\}$	$\{c\}$
$\{c, s\}$	$\{c\}$	$\{s\}$
$\{n, s\}$	$\{s\}$	$\{s\}$

Choice Replies of Alice at (γ_A, γ_B)

Bob

Alice

	L	M	R
U	${}^{A}n$	${}^{A}c$	n
M	c	s	${}^{A}c$
D	${}^{A}n$	s	${}^{A}s$

Choices	$C_A^{(\gamma_A,\gamma_B)}$	$C_B^{(\gamma_A,\gamma_B)}$
$\{c, n, s\}$	$\{c, s\}$	$\{n, s\}$
$\{c, n\}$	$\{n\}$	$\{c\}$
$\{c, s\}$	$\{c\}$	$\{s\}$
$\{n, s\}$	$\{s\}$	$\{s\}$

Choice Replies of Bob at (γ_A, γ_B)

Bob

Alice

	L	M	R
U	n	c^B	n
M	c	s^B	c
D	n	s^B	s^B

Choices	$C_A^{(\gamma_A,\gamma_B)}$	$C_B^{(\gamma_A,\gamma_B)}$
$\{c, n, s\}$	$\{c, s\}$	$\{n, s\}$
$\{c, n\}$	$\{n\}$	$\{c\}$
$\{c, s\}$	$\{c\}$	$\{s\}$
$\{n, s\}$	$\{s\}$	$\{s\}$

Nash Outcomes at (γ_A, γ_B)

$$\begin{array}{cc|ccc} & & & \text{Bob} & \\ & & L & M & R \\ \hline & U & {}^{A}n & {}^{A}\mathbf{c}^{B} & n \\ \text{Alice} & M & c & s^{B} & {}^{A}c \\ & D & {}^{A}n & s^{B} & {}^{A}\mathbf{s}^{B} \end{array}$$

Nash Equilibrium profiles at (γ_A, γ_B): $(U, M); (D, R)$

Nash Equilibrium outcomes at (γ_A, γ_B): $\boldsymbol{\{c, s\}}$

Choices	$C_A^{(\gamma_A,\gamma_B)}$	$C_B^{(\gamma_A,\gamma_B)}$
$\{c, n, s\}$	$\{c, s\}$	$\{n, s\}$
$\{c, n\}$	$\{n\}$	$\{c\}$
$\{c, s\}$	$\{c\}$	$\{s\}$
$\{n, s\}$	$\{s\}$	$\{s\}$

State by State Nash Equilibrium Outcomes

(ρ_A, ρ_B)

	L	M	R
U	**n**	c	**n**
M	c	**s**	c
D	n	**s**	s

N.Eq $= \{\boldsymbol{n}, \boldsymbol{s}\}$

(ρ_A, γ_B)

	L	M	R
U	**n**	c	**n**
M	c	**s**	c
D	n	**s**	s

N.Eq $= \{\boldsymbol{n}, \boldsymbol{s}\}$

(γ_A, ρ_B)

	L	M	R
U	n	**c**	n
M	c	s	c
D	**n**	s	s

N.Eq $= \{\boldsymbol{c}, \boldsymbol{n}\}$

(γ_A, γ_B)

	L	M	R
U	n	**c**	n
M	c	s	c
D	n	s	**s**

N.Eq $= \{\boldsymbol{c}, \boldsymbol{s}\}$

State by State BR-Optimal Outcomes

Choices	$C_A^{(\rho_A,\rho_B)}$	$C_B^{(\rho_A,\rho_B)}$
$\{c, n, s\}$	$\{n\}$	$\{s\}$
$\{c, n\}$	$\{n\}$	$\{n\}$
$\{c, s\}$	$\{c, s\}$	$\{s\}$
$\{n, s\}$	$\{n\}$	$\{s\}$

BR-Optima $= \boldsymbol{\{n, s\}}$

Choices	$C_A^{(\rho_A,\gamma_B)}$	$C_B^{(\rho_A,\gamma_B)}$
$\{c, n, s\}$	$\{n\}$	$\{n\}$
$\{c, n\}$	$\{n\}$	$\{n\}$
$\{c, s\}$	$\{s\}$	$\{s\}$
$\{n, s\}$	$\{n\}$	$\{s\}$

BR-Optima$= \boldsymbol{\{n, s\}}$

Choices	$C_A^{(\gamma_A,\rho_B)}$	$C_B^{(\gamma_A,\rho_B)}$
$\{c, n, s\}$	$\{n\}$	$\{c\}$
$\{c, n\}$	$\{n\}$	$\{c\}$
$\{c, s\}$	$\{c\}$	$\{c\}$
$\{n, s\}$	$\{n, s\}$	$\{n, s\}$

BR-Optima $= \boldsymbol{\{c, n\}}$

Choices	$C_A^{(\gamma_A,\gamma_B)}$	$C_B^{(\gamma_A,\gamma_B)}$
$\{c, n, s\}$	$\{c, s\}$	$\{n, s\}$
$\{c, n\}$	$\{n\}$	$\{c\}$
$\{c, s\}$	$\{c\}$	$\{s\}$
$\{n, s\}$	$\{s\}$	$\{s\}$

BR-Optima $= \boldsymbol{\{c, s\}}$

BR-Optimal Outcomes = Nash Outcomes

State by state, we have **BR-Optimal outcomes = Nash outcomes**:

(ρ_A, ρ_B) $\boldsymbol{\{n, s\}}$	(ρ_A, γ_B) $\boldsymbol{\{n, s\}}$
(γ_A, ρ_B) $\boldsymbol{\{c, n\}}$	(γ_A, γ_B) $\boldsymbol{\{c, s\}}$

This is nice! **But, how will they know the true state of the world**?

Ex-post Equilibrium

They **do not have to** if they are playing an **ex-post equilibrium**!

- The following are all the ex-post equilibria of $\mu = (M, g)$:
  - $\boldsymbol{\sigma}'^{*}$:
    $$\begin{array}{lll} \text{Alice} \rightarrow & \sigma_A^*(\rho_A) = U & \sigma_A^*(\gamma_A) = D \\ \text{Bob} \rightarrow & \sigma_B^*(\rho_B) = L & \sigma_B^*(\gamma_B) = R \end{array}$$
  - $\boldsymbol{\sigma}''^{*}$:
    $$\begin{array}{lll} \text{Alice} \rightarrow & \sigma_A'^*(\rho_A) = D & \sigma_A'^*(\gamma_A) = U \\ \text{Bob} \rightarrow & \sigma_B'^*(\rho_B) = M & \sigma_B'^*(\gamma_B) = M \end{array}$$
  - $\boldsymbol{\sigma}'''^{*}$:
    $$\begin{array}{lll} \text{Alice} \rightarrow & \sigma_A'^*(\rho_A) = M & \sigma_A'^*(\gamma_A) = U \\ \text{Bob} \rightarrow & \sigma_B'^*(\rho_B) = M & \sigma_B'^*(\gamma_B) = M \end{array}$$

    - $\boldsymbol{\sigma}''^{*}$ and $\boldsymbol{\sigma}'''^{*}$ are **outcome equivalent**.

Ex-post Equilibrium Outcomes

(ρ_A, ρ_B)

	L	M	R
U	**n**	c	**n**
M	c	**s**	c
D	n	**s**	s

$\boldsymbol{\sigma}'^{*}$, $\boldsymbol{\sigma}''^{*}$

EPEq $= \{$**n** , **s**$\}$

(ρ_A, γ_B)

	L	M	R
U	**n**	c	**n**
M	c	**s**	c
D	n	**s**	s

$\boldsymbol{\sigma}'^{*}$, $\boldsymbol{\sigma}''^{*}$

EPEq $= \{$**n** , **s**$\}$

(γ_A, ρ_B)

	L	M	R
U	n	**c**	n
M	c	s	c
D	**n**	s	s

$\boldsymbol{\sigma}'^{*}$, $\boldsymbol{\sigma}''^{*}$

EPEq $= \{$**n** , **c**$\}$

(γ_A, γ_B)

	L	M	R
U	n	**c**	n
M	c	s	c
D	n	s	**s**

$\boldsymbol{\sigma}'^{*}$, $\boldsymbol{\sigma}''^{*}$

EPEq $= \{$**s** , **c**$\}$

BR-Optimal outcomes = EPEq outcomes

State:	(ρ_A, ρ_B)	(ρ_A, γ_B)	(γ_A, ρ_B)	(γ_A, γ_B)
SCR: BR-Optimal	$\{n \quad , \quad s\}$	$\{n \quad , \quad s\}$	$\{c \quad , \quad n\}$	$\{c \quad , \quad s\}$
EPEq: $\sigma'^{*} \quad , \quad \sigma''^{*}$(or σ'''^{*})	$\{n \quad , \quad s\}$	$\{n \quad , \quad s\}$	$\{c \quad , \quad n\}$	$\{c \quad , \quad s\}$

BR-Optimal outcomes $=$ EPEq outcomes

$F = \{\mathbf{f}, \mathbf{f}'\}$ – Social Choice Set (BR-Optimal Outcomes)

$\mathbf{f}$	n	n	n	s
σ'^{*}	n	n	n	s
$\mathbf{f}'$	s	s	c	c
$\sigma''^{*} \equiv \sigma'''^{*}$	s	s	c	c

That is, we have

$$\begin{aligned} g(\sigma'^{*}(\theta)) &= \mathbf{f}(\theta) \text{ for all } \theta, \\ g(\sigma''^{*}(\theta)) &= \mathbf{f}'(\theta) \text{ for all } \theta. \\ g(\sigma'''^{*}(\theta)) &= \mathbf{f}'(\theta) \text{ for all } \theta. \end{aligned}$$

This is **full** ex-post implementation.

Why not Partial Implementation?: Revelation Principle fails!

The mechanism (M, g) **partially** implements the social choice function **f** in ex-post equilibrium as $\mathbf{f}(\theta) = g(\sigma'^{*}(\theta))$ for all $\theta \in \Theta$.

	(ρ_A, ρ_B)	(ρ_A, γ_B)	(γ_A, ρ_B)	(γ_A, γ_B)
f	n	n	n	s
$\boldsymbol{\sigma}^*$	n	n	n	s

Why not Partial Implementation?: Revelation Principle fails!

The corresponding **direct mechanism**:

Bob

		ρ_B	γ_B
Alice	ρ_A	n	n
	γ_A	n	s

$\mathbf{f} = g \circ \boldsymbol{\sigma}^*$

(ρ_A, ρ_B)	(ρ_A, γ_B)	(γ_A, ρ_B)	(γ_A, γ_B)
n	n	n	s

Why not Partial Implementation?: Revelation Principle fails!

The corresponding **direct mechanism** does **not** partially implement **f** in ex-post equilibrium:

Bob

		ρ_B	γ_B
Alice	ρ_A	n	n
	γ_A	n	s

This is because n is **not** chosen from $\{n, s\}$ by Alice at (ρ_A, γ_B).

Choices	$C_A^{(\rho_A,\gamma_B)}$	$C_B^{(\rho_A,\gamma_B)}$
$\{c, n, s\}$	$\{n\}$	$\{n\}$
$\{c, n\}$	$\{n\}$	$\{n\}$
$\{c, s\}$	$\{s\}$	$\{s\}$
$\{n, s\}$	$\{\mathbf{s}\}$	$\{s\}$

In this paper

We provide necessary as well as sufficient conditions for behavioral implementation under incomplete information.

- In doing so, we restrict our attention to
  - **full** implementation
    - the set of equilibrium outcomes fully coincide with a predetermined social goal,
  - **ex-post** equilibrium
    - measurable strategies that induce a Nash equilibrium at every state of the world.

Why full implementation?

- The power of **partial** implementation relies heavily on **the revelation principle**.

> [*The direct revelation partial implementation*] *does assure that the resulting outcome will be an equilibrium of some game; however, there may be others as well. This problem is sometimes dismissed with an argument that as long as truthful revelation is an equilibrium, it will somehow be the salient equilibrium even if there are other equilibria. (Postlewaite and Schmeidler, 1986).*

Why full implementation?

- In our setup, **the revelation principle** does **not** hold!
  - The salience of a truth-telling equilibrium is **not** reasonable.
- We **cannot** restrict attention to direct revelation mechanisms without loss of generality if individual choices are not rational.
- Identifying mechanisms for **full implementation** are also useful as full implementation **implies** partial implementation.

Why ex-post equilibrium?

Ex-post equilibrium is **plausible** in our setup since:

- Expected Utility hypothesis fails due to lack of rationality (in particular completeness and transitivity).
  - Hence, Bayesian Nash Equilibrium is not plausible in our setup.
- No common prior assumption is required.
- No complex (Bayesian) updating is required.
- No-regret is obtained – at least when individuals are rational.
- Convenient for handling interdependence.
- Robust to informational assumptions about the environment.

Preliminaries I

The **environment** can be summarized as $\langle N, X, \Theta, C_i^\theta{}_{\{i \in N, \theta \in \Theta\}} \rangle$.

- $N = \{1, ..., n\}$ set of **players**,
- X denotes the set of all possible **alternatives**
    - $\mathcal{X}$ denotes the set of all **subsets** of X,
- $\Theta = \times_{i \in N} \Theta_i$, denotes the set of all possible **states** of the world,
    - θ_i denotes the **private information** of i.
- $C_i^\theta : \mathcal{X} \rightarrow \mathcal{X}$ describes the **choice** behavior of agent i at θ.
    - $C_i^\theta(S) \subseteq S$, for all $S \in \mathcal{X}$.

Preliminaries II

- The social choice rule is described as a Social Choice Set, F.
$$F \subset \{f | f : \Theta \to X\}$$

- $\mu = (M, g)$ denotes a **mechanism** where
  - $M_i \neq \emptyset$ is the **message space** of agent $i \in N$ and
  - $g : M \to X$ is the **outcome function**. $(M \equiv \times_{i \in N} M_i.)$

- $\sigma_i : \Theta_i \to M_i$ denotes a **strategy** of agent i under the mechanism μ.

- The **opportunity set** of agent i under μ for each m_{-i} is given by
$$O_i^{\mu}(m_{-i}) = \{g(m_i, m_{-i}) \in X : m_i \in M_i\}.$$

Ex-post Equilibrium of a Mechanism

Definition

A strategy profile $\sigma^* : \Theta \to M$ is an **ex-post equilibrium** of $\mu = (M, g)$ if for **every** $\theta \in \Theta$ we have

$$g(\sigma_i^*(\theta_i), \sigma_{-i}^*(\theta_{-i})) \in C_i^{(\theta_i, \theta_{-i})}(O_i^\mu(\sigma_{-i}^*(\theta_{-i}))) \qquad \forall i \in N.$$

Ex-post Implementation

Definition

A social choice **set** F is is said to be **ex-post implementable** if there exists a mechanism $\mu = (M, g)$ such that:

(i) For every $f \in F$, there exists an ex-post equilibrium σ^* of $\mu = (M, g)$ that satisfies

$$[f = g \circ \sigma^*], \text{ i.e., } f(\theta) = g(\sigma^*(\theta)) \text{ for all } \theta \in \Theta;$$

(ii) For every ex post equilibrium σ^* of $\mu = (M, g)$, there exists $f \in F$ such that:

$$[g \circ \sigma^* = f], \text{ i.e., } g(\sigma^*(\theta)) = f(\theta) \text{ for all } \theta \in \Theta.$$

Deception and Deception Profile

- $\alpha_i : \Theta_i \to \Theta_i$ denotes a possible **deception** by agent $i \in N$.
  - $\alpha_i(\theta_i)$ can be interpreted as i's reported type.
- $\alpha(\theta) = (\alpha_1(\theta_1), \alpha_2(\theta_2), \ldots, \alpha_n(\theta_n))$. denotes a **profile of deceptions**.

Necessity

Consistent Collections of Sets under Incomplete Information

Definition

A collection of sets $\mathbb{S} := \{S_i(f, \theta_{-i}) | i \in N, f \in F, \theta_{-i} \in \Theta_{-i}\} \subset \mathcal{X}$ is **consistent with the SCS $F \in \mathcal{F}$ under incomplete information** if for every SCF $f \in F$, we have

(i) for all $i \in N$, $f(\theta_i', \theta_{-i}) \in C_i^{(\theta_i', \theta_{-i})}(S_i(f, \theta_{-i}))$ for each $\theta_i' \in \Theta_i$, and

(ii) for any deception profile α with $f \circ \alpha \notin F$, there exists $\theta^* \in \Theta$ and $i^* \in N$ such that $f(\alpha(\theta^*)) \notin C_{i^*}^{\theta^*}(S_{i^*}(f, \alpha_{-i^*}(\theta^*_{-i^*})))$.

Consistency under Incomplete Information

- A collection of sets $\mathbb{S}$ is **consistent with an SCS F under incomplete information** if
    - i. Given any $i \in N$, any $f \in F$, and any $\theta_{-i} \in \Theta_{-i}$, it must be that when i's type is θ'_i, his choices from $S_i(f, \theta_{-i})$ at state (θ'_i, θ_{-i}) contains $f(\theta'_i, \theta_{-i})$ for all $\theta'_i \in \Theta_i$; and
    - ii. given any $f \in F$, whenever there is a **deception** profile α that leads to an outcome **not** compatible with the SCS F, ($f \circ \alpha \notin F$), there exists an **informant** state θ^* and an **informant** individual i^* such that
        - i^* does **not** choose at state θ^* the alternative $f(\alpha(\theta^*))$ from $S_{i^*}(f, \alpha_{-i^*}(\theta^*_{-i^*}))$.

Necessity

Theorem

If an SCS F is ex-post implementable, then there exists a collection of sets $\mathbb{S} := \{S_i(f, \theta_{-i}) | i \in N, f \in F, \theta_{-i} \in \Theta_{-i}\}$ *consistent with F under incomplete information.*

Ex-post Choice Monotonicity

Definition

F is **ex-post-choice-monotonic** if, for every $f \in F$ and deception profile α with $f \circ \alpha \notin F$, there is $\theta^* \in \Theta, i^* \in N, S^* \in \mathcal{X}$ such that

$$\text{(i)} \quad f(\alpha(\theta^*)) \notin C_{i^*}^{\theta^*}(S^*),$$

$$\text{(ii)} \quad f((\theta'_{i^*}, \alpha_{-i^*}(\theta^*_{-i^*}))) \in C_{i^*}^{(\theta'_{i^*}, \alpha_{-i^*}(\theta^*_{-i^*}))}(S^*) \text{ for all } \theta'_{i^*} \in \Theta_{i^*}.$$

- Consistency under Incomplete Information implies Ex-post Choice Monotonicity.

Proposition

If F is ex-post implementable, then F is ex-post choice monotonic.

What is Ex-post Choice Monotonicity?

Ex-post Choice-Monotonicity requires that when there is an attempt of deception that would lead to a non-optimal outcome, there must exist

- an **informant** state,
- an **informant/whistle-blower** agent for this state, and
- an **informant/reward** set for this whistle-blower,

such that

- (i) the whistle-blower would not choose the outcome arising due to the undesirable deception from the reward set in the informant state;
- (ii) the whistle-blower does not have an incentive to falsely alert the designer when the outcome is optimal, i.e., compatible with the social choice set in question.

Quasi-Ex-post Choice Incentive Compatibility

Definition

F is **quasi-ex-post choice incentive compatible** if, for every $f \in F$, $\theta \in \Theta$, $i \in N$ there exists $S \in \mathcal{X}$ such that

i) $S \supseteq \{f(\theta_i', \theta_{-i}) | \theta_i' \in \Theta_i\}$,

ii) $f(\theta) \in C_i^\theta(S)$.

- Consistency under Incomplete Information implies Quasi-Ex-post Choice Incentive Compatibility.

Proposition

If F is ex-post implementable, then F is quasi-ex-post choice incentive compatible.

What is Quasi-Ex-post Choice Incentive Compatibility?

- Quasi-ex-post choice incentive compatibility condition describes a necessary condition for partial ex-post implementation of any $f \in F$.

- A necessary and sufficient condition for revelation principle (direct revelation partial ex-post implementation) of any $f \in F$ is
  - For every $\theta \in \Theta$, $i \in N$, $f(\theta) \in C_i^{\theta}(\{f(\theta_i', \theta_{-i}) | \theta_i' \in \Theta_i\})$.

IIA is sufficient for Revelation Principle

Proposition

If individual choices satisfy the IIA, then quasi-ex-post choice incentive compatibility implies the revelation principle.

- Put differently, the **revelation principle** holds whenever individuals' choices satisfy the **IIA**!

Sufficiency results with three or more individuals

Choice Incompatible Pair Property

Definition

S satisfies the choice incompatible pair property at θ if for each $x \in S$ there exist $i, j \in N$ such that $x \notin C_i^{\theta}(S)$ **and** $x \notin C_j^{\theta}(S)$.

- A set satisfies the choice incompatible pair property at a particular state if for each alternative in this set, there exists a pair of individuals who do not choose this alternative from this set at the particular state.
- This guarantess any alternative in this set can be chosen by at most $n-2$ individuals at the particular state.

Choice Incompatible Pair Property

Definition

We say that S satisfies the choice incompatible pair property at θ if for each $x \in S$ there exist $i, j \in N$ s.t. $x \notin C_i^{\theta}(S)$ **and** $x \notin C_j^{\theta}(S)$.

- The choice incompatible pair property is similar to the economic environment assumption of the rational domain.
    - Yet, it is much weaker as it is defined for a set.

Sufficieny with Choice Incompatible Pair Property

Theorem

Let $n \geq 3$. If F is an SCS for which there exist

(*i*) *a collection of sets $\mathbb{S} := \{S_i(f, \theta_{-i}) : i \in N, f \in F, \theta_{-i} \in \Theta_{-i}\}$ consistent with F under incomplete information, and*

(*ii*) *a set of alternatives $\bar{X} \subseteq X$ with $\bigcup_{S \in \mathbb{S}} S \subseteq \bar{X}$ which satisfies the choice incompatible pair property at every state $\theta \in \Theta$,*

then F is **ex-post implementable**.

Sufficieny with Choice Incompatible Pair Property

- When there are three or more individuals an SCS F is **ex-post implementable** whenever
  - (i) there exists a collection of sets $\mathbb{S}$ consistent with F under incomplete information, and
  - (ii) there exists a set of alternatives $\bar{X}$ which contains every alternative in $\mathbb{S}$ and satisfies the choice incompatible pair property at every state of the world.

Choice No-Veto-Power Property

Definition

A social choice function f satisfies **choice no-veto-power property on S at θ** if $x \in C_i^{\theta}(S)$ for all $i \in N \setminus \{j\}$ implies $f(\theta) = x$.

- The choice-no-veto power property of f on S at θ requires that if a particular alternative is chosen from S by at least $n-1$ individuals at θ, then this particular alternative must be f-optimal at θ.
    - We note that it is weaker than its analog in the rational domain since it is defined on a particular set.

Consistency No-Veto-Power Property

Definition

An SCS F satisfies the **consistency-no-veto property** whenever there exist

- (i) a collection of sets $\mathbb{S} := \{S_i(f, \theta_{-i}) : i \in N, f \in F, \theta_{-i} \in \Theta_{-i}\}$ such that for all $f \in F$ and for all $i \in N$, $f(\theta'_i, \theta_{-i}) \in C_i^{(\theta'_i, \theta_{-i})}(S_i(f, \theta_{-i}))$ for each $\theta'_i \in \Theta_i$,
- (ii) and a set of alternatives $\bar{X} \subseteq X$ with $\bigcup_{S \in \mathbb{S}} S \subseteq \bar{X}$

such that for any **collection of product sets** of states $\{\bar{\Theta}_f\}_{f\ \in F}$ with $\bar{\Theta} = \bigcup_{f \in F} \bar{\Theta}_f \subset \Theta$, there exists $f^* \in F$ such that

- (iii) f^* satisfies choice no-veto-power property on $\bar{X}$ at every $\theta \in \Theta \setminus \bar{\Theta}$, and
- (iv) if for any $f \in F$ and any deception profile α, $f(\alpha(\theta)) \neq f^*(\theta)$ for some $\theta \in \bar{\Theta}_f$, then there exists $i^* \in N$ and $\theta^* \in \bar{\Theta}_f$ such that $f(\alpha(\theta^*)) \notin C_{i^*}^{\theta^*}(S_{i^*}(f, \alpha_{-i^*}(\theta^*_{-i^*})))$.

Consistency No-Veto-Power Property is sufficient for ex-post implementation

Theorem

Let $n \geq 3$. If an SCS F satisfies the consistency-no-veto property, then F is ex-post implementable.

Consistency No-Veto-Power Property

Given an SCS F, the consistency-no-veto property requires the existence of a collection of sets $\mathbb{S}$ and a set of alternatives $\bar{X}$ which contains every alternative that appears in $\mathbb{S}$ such that:

- Given any $i \in N$, any $f \in F$, and any $\theta_{-i} \in \Theta_{-i}$, it must be that when i's type is θ'_i, his choices from $S_i(f, \theta_{-i})$ at state (θ'_i, θ_{-i}) contains $f(\theta'_i, \theta_{-i})$ for all $\theta'_i \in \Theta_i$; and
- for any collection of product sets of states $\{\bar{\Theta}_f\}_{f \in F}$ with $\bar{\Theta} = \bigcup_{f \in F} \bar{\Theta}_f \subset \Theta$, there is an SCF f^* in F such that
  - if $\theta \in \Theta \setminus \bar{\Theta}$, then f^* obeys the choice no-veto-power property on $\bar{X}$ at θ, and
  - if a deception profile α and an SCF $f \in F$ lead to an outcome different than $f^*(\theta)$ for some $\theta \in \bar{\Theta}_f$, then there exists a whistle-blower $i^* \in N$ and an informant state θ^* such that i^* does **not** choose $f(\alpha(\theta^*))$ from $S_{i^*}(f, \alpha_{-i^*}(\theta^*_{-i^*}))$ at θ^*.

Corollary for a social choice function

Corollary

Let $n \geq 3$*. An* **SCF** $f : \Theta \to X$ *is* **ex-post implementable** *if there exists a collection* $\mathbb{S} := \{S_i(f, \theta_{-i}) : i \in N, \theta_{-i} \in \Theta_{-i}\}$ *s.t.* $f(\theta'_i, \theta_{-i}) \in C_i^{(\theta'_i, \theta_{-i})}(S_i(f, \theta_{-i}))$ *for each* $\theta'_i \in \Theta_i$ *and there exists* $\bar{X} \subseteq X$ *with* $\bigcup_{S \in \mathbb{S}} S \subseteq \bar{X}$ *s.t. for any product set* $\bar{\Theta} \subset \Theta$,

- (i) f *satisfies* **choice no-veto-power property** *on* $\bar{X}$ *at every* $\theta \in \Theta \setminus \bar{\Theta}$*, and*
- (ii) *for any* **deception** *profile* α *with* $f(\alpha(\theta)) \neq f(\theta)$ *for some* $\theta \in \bar{\Theta}$*, there exists* $i^* \in N$ *and* $\theta^* \in \bar{\Theta}$ *such that* $f(\alpha(\theta^*)) \notin C_{i^*}^{\theta^*}(S_{i^*}(f, \alpha_{-i^*}(\theta^*_{-i^*})))$.

The mechanism

$\mu = (M, g)$: $M_i = F \times \Theta_i \times \bar{X} \times N$ and $g : \mathcal{M} \to X$ is as follows:

Rule 1 : $g(m) = f(\theta)$ if $m_i = (f, \theta_i, \cdot, \cdot)$ for all $i \in N$,

Rule 2 : $g(m) = \begin{cases} x_j & \text{if } x_j \in S_j(f, \theta_{-j}), \\ \bar{x}(j, f, \theta_{-j}) & \text{otherwise.} \end{cases}$ if $m_i = (f, \theta_i, \cdot, \cdot)$ for all $i \in N \setminus \{j\}$ and $m_j = (\tilde{f}, \tilde{\theta}_j, x_j, \cdot)$ with $\tilde{f} \neq f$,

Rule 3 : $g(m) = x_j$ where $j = \sum_i k_i$ (mod n) otherwise.

where $\bar{x}(j, f, \theta_{-j})$ is an arbitrary element from $S_j(f, \theta_{-j})$.

Two Individuals

Two-Individual Consistency

Definition

$\mathbb{S}_1 := \{S_1(f,\theta_2) | f \in F, \theta_2 \in \Theta_2\}$ and $\mathbb{S}_2 := \{S_2(f,\theta_1) | f \in F, \theta_1 \in \Theta_1\}$ are **two-individual consistent with the SCS F under incomplete info** if

(*i*) for all $f \in F$, $f(\theta_1', \theta_2) \in C_1^{(\theta_1',\theta_2)}(S_1(f,\theta_2))$ for each $\theta_1' \in \Theta_1$,

(*ii*) for all $f \in F$, $f(\theta_1, \theta_2') \in C_2^{(\theta_1,\theta_2')}(S_2(f,\theta_1))$ for each $\theta_2' \in \Theta_2$,

(*iii*) for all $f, f' \in F$, $S_1(f,\theta_2) \cap S_2(f',\theta_1) \neq \emptyset$ for each $\theta_1 \in \Theta_1$ and $\theta_2 \in \Theta_2$,

(*iv*) for all $f \in F$, if $f \circ \alpha \notin F$, then there exists $\theta^* = (\theta_1^*, \theta_2^*) \in \Theta$ s.t. $f(\alpha(\theta^*)) \notin C_1^{\theta^*}(S_1(f, \alpha_2(\theta_2^*)))$ **or** $f(\alpha(\theta^*)) \notin C_2^{\theta^*}(S_2(f, \alpha_1(\theta_1^*)))$.

- This is slightly **stronger** than consistency from before.
  - The only new condition is (iii)!

Necessity: Two Individuals

Theorem

Let $n = 2$*. If an SCS F is ex-post implementable, then* **there exist** *collections of sets* $\mathbb{S}_1 := \{S_1(f, \theta_2) | f \in F, \theta_2 \in \Theta_2\}$ *and* $\mathbb{S}_2 := \{S_2(f, \theta_1) | f \in F, \theta_1 \in \Theta_1\}$ *that are two-individual consistent with F under incomplete information.*

Practicality – the Motivating Example

Let's investigate the collections that satisfy **two-individual consistency** in our motivating example:

- $\mathbb{S}_A = \{S_A(f, \rho_B), S_A(f, \gamma_B), S_A(f', \rho_B), S_A(f', \gamma_B)\}$,
- $\mathbb{S}_B = \{S_B(f, \rho_A), S_B(f, \gamma_A), S_B(f', \rho_A), S_B(f', \gamma_A)\}$.

Two-individual consistency for Alice

(i) of two-individual consistency implies (for Alice):

- $\underline{S_A(f,\rho_B)}$: $f(\rho_A,\rho_B)=n$ and $f(\gamma_A,\rho_B)=n$ imply $n\in C_A^{(\rho_A,\rho_B)}(S_A(f,\rho_B))$ and $n\in C_A^{(\gamma_A,\rho_B)}(S_A(f,\rho_B))$. There are **four** such sets: $\{c,n,s\},\{c,n\},\{n,s\},\{n\}$.
- $\underline{S_A(f,\gamma_B)}$: $f(\rho_A,\gamma_B)=n$ and $f(\gamma_A,\gamma_B)=s$ imply $n\in C_A^{(\rho_A,\gamma_B)}(S_A(f,\gamma_B))$ and $s\in C_A^{(\gamma_A,\gamma_B)}(S_A(f,\rho_B))$. There is only **one** such set: $\{c,n,s\}$.
- $\underline{S_A(f',\rho_B)}$: $f'(\rho_A,\rho_B)=s$ and $f'(\gamma_A,\rho_B)=c$ imply $s\in C_A^{(\rho_A,\rho_B)}(S_A(f',\rho_B))$ and $c\in C_A^{(\gamma_A,\rho_B)}(S_A(f',\rho_B))$. There is only **one** such set as well: $\{c,s\}$.
- $\underline{S_A(f',\gamma_B)}$: $f'(\rho_A,\gamma_B)=s$ and $f'(\gamma_A,\gamma_B)=c$ imply $s\in C_A^{(\rho_A,\gamma_B)}(S_A(f',\rho_B))$ and $c\in C_A^{(\gamma_A,\gamma_B)}(S_A(f',\rho_B))$. There is again only **one** such set: $\{c,s\}$.

Two-individual consistency for Bob

(ii) of two-individual consistency implies (for Bob):

- $\underline{S_B(f,\rho_A)}$: $f(\rho_A,\rho_B)=n$ and $f(\rho_A,\gamma_B)=n$ imply $n\in C_B^{(\rho_A,\rho_B)}(S_B(f,\rho_A))$ and $n\in C_B^{(\rho_A,\gamma_B)}(S_B(f,\rho_A))$. There are **two** such sets: $\{c,n\}$ and $\{n\}$.
- $\underline{S_B(f,\gamma_A)}$: $f(\gamma_A,\rho_B)=n$ and $f(\gamma_A,\gamma_B)=s$ imply $n\in C_B^{(\gamma_A,\rho_B)}(S_B(f,\gamma_A))$ and $s\in C_B^{(\gamma_A,\gamma_B)}(S_B(f,\gamma_A))$. There is only **one** such set: $\{n,s\}$.
- $\underline{S_B(f',\rho_A)}$: $f'(\rho_A,\rho_B)=s$ and $f'(\rho_A,\gamma_B)=s$ imply $s\in C_B^{(\rho_A,\rho_B)}(S_B(f',\rho_A))$ and $s\in C_B^{(\rho_A,\gamma_B)}(S_B(f',\rho_A))$. There are **three** such sets $\{c,s\}$ and $\{n,s\}$ and $\{s\}$.
- $\underline{S_B(f',\gamma_A)}$: $f'(\gamma_A,\rho_B)=c$ and $f'(\gamma_A,\gamma_B)=c$ imply $c\in C_B^{(\gamma_A,\rho_B)}(S_B(f',\gamma_A))$ and $c\in C_B^{(\gamma_A,\gamma_B)}(S_B(f',\gamma_A))$. There are **two** such sets $\{c,n\}$ and $\{c\}$.

Implications of (*iii*) of two-individual consistency

(*iii*) of two-individual consistency:

- $S_A(f, \theta_B) \cap S_B(f', \theta_A) \neq \emptyset$ and $S_A(f', \theta_B) \cap S_B(f, \theta_A) \neq \emptyset$ for each $\theta_A \in \{\rho_A, \gamma_A\}$ and $\theta_B \in \{\rho_B, \gamma_B\}$.

In particular, $S_A(f', \rho_B) \cap S_B(f, \rho_A) \neq \emptyset \implies S_B(f, \rho_A) = \{c, n\}$.

Two-individual consistency (i), (ii), (iii) uniquely identifies 5 out of 8 sets

- $S_A(f, \gamma_B) = \{c, n, s\}$; $S_A(f', \rho_B) = \{c, s\}$; $S_A(f', \gamma_B) = \{c, s\}$,
- $S_B(f, \gamma_A) = \{n, s\}$; $S_B(f, \rho_A) = \{c, n\}$.

$\mathbb{S}_A = \{S_A(f, \rho_B), \{c, n, s\}, \{c, s\}\}$;
$\mathbb{S}_B = \{\{c, n\}, \{n, s\}, S_B(f', \rho_A), S_B(f', \gamma_A)\}$.

- It is possible to narrow down $\mathbb{S}_A$ and $\mathbb{S}_B$ further by employing condition (iv) of two-individual consistency. (tedious by hand!)
- There is **not** a unique pair of collections that satisfy two-individual consistency—as we shall see!

Sufficiency results with two individuals

Choice Incompatibility

Definition

F involves **choice incompatibility** among a set of alternatives $S \in \mathcal{X}$ and collections of sets $\mathbb{S}_1 := \{S_1(f, \theta_2) | f \in F, \theta_2 \in \Theta_2\} \subset \mathcal{X}$ and $\mathbb{S}_2 := \{S_2(f, \theta_1) | f \in F, \theta_1 \in \Theta_1\} \subset \mathcal{X}$ at θ if

- (*i*) $x \in C_i^\theta(S)$ implies $x \notin C_j^\theta(S)$, $i \neq j$; and
- (*ii*) for any $T \in \mathbb{S}_i$, $x \in C_i^\theta(T)$ implies $x \notin C_j^\theta(S)$, $i = 1, 2$ and $i \neq j$; and
- (*iii*) for any deception profile α and any $f, f' \in F$ with $f \neq f'$, $x \in C_i^\theta(S_i(f, \alpha_j(\theta_j)))$ implies $x \notin C_j^\theta(S_j(f', \alpha_i(\theta_i)))$, $i = 1, 2$ and $i \neq j$.

Choice Incompatibility

- F involves choice incompatibility among a non-empty set of alternatives S and non-empty collections of sets $\mathbb{S}_1$ and $\mathbb{S}_2$ at state θ means that the individual choices at θ are **not** aligned when
  i. both individuals make choices separately from S; and
  ii. one individual, i, is making a choice from a set in $\mathbb{S}_i$ and the other individual, j, is making a choice from S where $i, j = 1, 2$ with $i \neq j$; and
  iii. individual i makes a choice from a set in $\mathbb{S}_i$ that is associated with a particular SCF f and the other individual, j, makes a choice from a set in $\mathbb{S}_j$ which is associated with a different SCF $f' \neq f$ while $f, f' \in F$ and $i, j = 1, 2$ with $i \neq j$.

That is, choice incompatibility conditions require that there is sufficient disagreement between the two individuals at a given state of the world.

Sufficiency with Choice Incompatibility

Theorem

Let $n = 2$. If F is an SCS for which there exist

(*i*) *collections of sets $\mathbb{S}_1 := \{S_1(f, \theta_2) | f \in F, \theta_2 \in \Theta_2\}$ and $\mathbb{S}_2 := \{S_2(f, \theta_1) | f \in F, \theta_1 \in \Theta_1\}$ which are two-individual consistent with F under incomplete information, and*

(*ii*) *a set of alternatives $\bar{X} \subseteq X$ with $\bigcup_{S \in \mathbb{S}_1 \cup \mathbb{S}_2} S \subseteq \bar{X}$ such that F involves choice incompatibility among $\bar{X}$ and $\mathbb{S}_1$ and $\mathbb{S}_2$ at every $\theta \in \Theta$,*

then F is **ex-post implementable***.*

Sufficiency with Choice Incompatibility

- When there are two individuals, F is ex-post implementable whenever
    - (i) there exist collections of sets $\mathbb{S}_1$ and $\mathbb{S}_2$ that are **two-individual consistent with F under incomplete information**, and
    - (ii) there exists a set of alternatives $\bar{X}$ that **contains** every alternative in $\mathbb{S}_1$ and $\mathbb{S}_2$; and **choice incompatibility** among $\bar{X}$ and $\mathbb{S}_1$ and $\mathbb{S}_2$ hold at **every state of the world**.

Choice Unanimity

Definition

F respects **choice unanimity** on a non-empty set of alternatives $S \in \mathcal{X}$ and non-empty collections of sets $\mathbb{S}_1 := \{S_1(f, \theta_2) | f \in F, \theta_2 \in \Theta_2\} \subset \mathcal{X}$ and $\mathbb{S}_2 := \{S_2(f, \theta_1) | f \in F, \theta_1 \in \Theta_1\} \subset \mathcal{X}$ at θ if there exists $f^* \in F$ s.t.

(*i*) $x \in C_1^\theta(S) \cap C_2^\theta(S)$ implies $f^*(\theta) = x$; and

(*ii*) for any $T \in \mathbb{S}_i$, $x \in C_i^\theta(T) \cap C_j^\theta(S)$ implies $f^*(\theta) = x$, for $i, j = 1, 2$ with $i \neq j$; and

(*iii*) for any deception profile α and $f, f' \in F$ with $f \neq f'$, $x \in C_i^\theta(S_i(f, \alpha_j(\theta_j))) \cap C_j^\theta(S_j(f', \alpha_i(\theta_i)))$ implies $f^*(\theta) = x$.

Consistency-Unanimity

Definition

Let $n = 2$. F satisfies the **consistency-unanimity property** whenever there exist collections $\mathbb{S}_1 := \{S_1(f, \theta_2) | f \in F, \theta_2 \in \Theta_2\} \subset \mathcal{X}$ and $\mathbb{S}_2 := \{S_2(f, \theta_1) | f \in F, \theta_1 \in \Theta_1\} \subset \mathcal{X}$ s.t

- $(i - ii)$ for all $f \in F$, $f(\theta_i', \theta_j) \in C_i^{(\theta_i', \theta_j)}(S_i(f, \theta_j))$ for each $\theta_i' \in \Theta_i$, and
- (iii) for all $f, f' \in F$, $S_1(f, \theta_2) \cap S_2(f', \theta_1) \neq \emptyset$ for each $\theta_1 \in \Theta_1$ and $\theta_2 \in \Theta_2$,

and there is a set of alternatives $\bar{X} \subseteq X$ with $\bigcup_{S \in \mathbb{S}_1 \cup \ \mathbb{S}_2} S \subseteq \bar{X}$ such that for any collection of product sets $\{\bar{\Theta}_f\}_{f \ \in F}$ with $\bar{\Theta} = \bigcup_{f \in F} \bar{\Theta}_f \subset \Theta$,

- (iv) F respects choice unanimity on $\bar{X}$ and $\mathbb{S}_1$ and $\mathbb{S}_2$ at every $\theta \in \Theta \setminus \bar{\Theta}$, and
- (v) for all $f \in F$ and deception profile α, if $f(\alpha(\theta)) \neq f^*(\theta)$ for some $\theta \in \bar{\Theta}_f$ where f^* is the SCF that satisfies (i)–(iii) of choice unanimity, then there exists $\theta^* \in \bar{\Theta}_f$ such that either $f(\alpha(\theta^*)) \notin C_1^{\theta^*}(S_1(f, \alpha_2(\theta_2^*)))$ or $f(\alpha(\theta^*)) \notin C_2^{\theta^*}(S_2(f, \alpha_1(\theta_1^*)))$.

Consistency-Unanimity is sufficient
$(n = 2)$

Theorem

Let $n = 2$. If an SCS F satisfies the consistency-unanimity property, then F is ex-post implementable.

Consistency-Unanimity in words

- Given a non-empty SCS F, there exist collections of sets $\mathbb{S}_i$ $(i = 1, 2)$ and a set of alternatives $\bar{X}$ which contains every alternative in $\mathbb{S}_1 \cup \mathbb{S}_2$ such that the following hold:
  - when i's type is θ_i' his choice from $S_i(f, \theta_j)$ at state (θ_i', θ_j) contains $f(\theta_i', \theta_j)$ for all $\theta_i' \in \Theta_i$, with $j = 1, 2$ and $i \neq j$;
  - any set in $\mathbb{S}_1$ must have a common element with any set in $\mathbb{S}_2$;
  - for any collection of product sets of states $\{\bar{\Theta}_f\}_{f\ \in F}$ with $\bar{\Theta} = \bigcup_{f \in F} \bar{\Theta}_f \subset \Theta$, there is an SCF f^* in F s.t.
    - F respects choice unanimity on $\bar{X}$ and $\mathbb{S}_1$ and $\mathbb{S}_2$ whenever $\theta \in \Theta \setminus \bar{\Theta}$; and
    - for any deception profile α and SCF $f \in F$ that lead to an outcome different than $f^*(\theta)$ for some $\theta \in \bar{\Theta}_f$, there exists a whistle-blower $i^* \in \{1, 2\}$ and an informant state $\theta^* \in \bar{\Theta}_f$ such that i^* does **not** choose $f(\alpha(\theta^*))$ from $S_{i^*}(f, \alpha_j(\theta_j^*))$ at θ^* where $j \in \{1, 2\}$ and $j \neq i^*$.

The mechanism

$\mu = (M, g)$: $M_i = \{0,1\} \times F \times \Theta_i \times \bar{S}_i \times \{0,1\}$ and $g : \mathcal{M} \to X$ is:

Rule 1 : $g(m) = f(\theta)$ if $m_i = (0, f, \theta_i, \cdot, \cdot)$ for both $i \in \{1,2\}$,

Rule 2.1 : $$g(m) = \begin{cases} x_1 & \text{if } x_1 \in S_1(f_2, \theta_2) \\ \bar{x}(1, f_2, \theta_2) & \text{otherwise,} \end{cases}$$ if $m_1 = (1, f_1, \theta_1, x_1, \cdot)$ and $m_2 = (0, f_2, \theta_2, x_2, \cdot)$,

Rule 2.2 : $$g(m) = \begin{cases} x_2 & \text{if } x_2 \in S_2(f_1, \theta_1) \\ \bar{x}(2, f_1, \theta_1) & \text{otherwise,} \end{cases}$$ if $m_1 = (0, f_1, \theta_1, x_1, \cdot)$ and $m_2 = (1, f_2, \theta_2, x_2, \cdot)$,

Rule 3 : $g(m) = \bar{x}(f_1, f_2, \theta_1, \theta_2)$ if $m_1 = (0, f_1, \theta_1, x_1, k_1)$ and $m_2 = (0, f_2, \theta_2, x_2, k_2)$ with $f_1 \neq f_2$,

Rule 4 : $g(m) = x_j$ if $m_1 = (1, f_1, \theta_1, x_1, k_1)$ and $m_2 = (1, f_2, \theta_2, x_2, k_2)$ with $$j = \begin{cases} 1 & \text{if } k_1 + k_2 \text{ is odd,} \\ 2 & \text{if } k_1 + k_2 \text{ is even.} \end{cases}$$

where $\bar{x}(j, f, \theta_{-j}) \in S_j(f, \theta_{-j})$ and $\bar{x}(f, f', \theta_1, \theta_2) \in S_1(f, \theta_2) \cap S_2(f', \theta_1)$.

Corollary for a social choice function

Corollary

Let $n = 2$. An SCF $f : \Theta \to X$ is ex-post implementable whenever there are collections of sets $\mathbb{S}_i := \{S_i(f, \theta_j) | \theta_j \in \Theta_j\} \subset \mathcal{X}$ with $i, j = 1, 2$ and $i \neq j$ such that

(*i*) $f(\theta_1', \theta_2) \in C_1^{(\theta_1', \theta_2)}(S_1(f, \theta_2))$ *for each* $\theta_1' \in \Theta_1$, *and*
$f(\theta_1, \theta_2') \in C_2^{(\theta_1, \theta_2')}(S_2(f, \theta_1))$ *for each* $\theta_2' \in \Theta_2$, *and*
$S_1(f, \theta_2) \cap S_2(f, \theta_1) \neq \emptyset$ *for each* $\theta_1 \in \Theta_1$ *and* $\theta_2 \in \Theta_2$,

and there is a set of alternatives $\bar{X} \subseteq X$ *with* $\bigcup_{S \in \mathbb{S}_1 \cup \ \mathbb{S}_2} S \subseteq \bar{X}$ *such that for any product set* $\bar{\Theta} \subseteq \Theta$,

(*ii*) $x \in C_1^{\theta}(\bar{X}) \cap C_2^{\theta}(\bar{X})$ *implies* $f(\theta) = x$, $x \in C_1^{\theta}(T) \cap C_2^{\theta}(\bar{X})$ *with* $T \in \mathbb{S}_1$ *implies* $f(\theta) = x$, *and* $x \in C_1^{\theta}(\bar{X}) \cap C_2^{\theta}(T')$ *with* $T' \in \mathbb{S}_2$ *implies* $f(\theta) = x$, *for each* $\theta \in \Theta \setminus \bar{\Theta}$, *and*

(*iii*) *for any deception profile* α, *if* $f(\alpha(\theta)) \neq f(\theta)$ *for some* $\theta \in \bar{\Theta}$, *then there exists* $\theta^* \in \bar{\Theta}$ *such that either* $f(\alpha(\theta^*)) \notin C_1^{\theta^*}(S_1(f, \alpha_2(\theta_2^*)))$ **or** $f(\alpha(\theta^*)) \notin C_2^{\theta^*}(S_2(f, \alpha_1(\theta_1^*)))$.

Application of the Corollary:

The following example is inspired from Masatlioglu and OK (2014), which presents a "*model of individual decision making when the endowment of an agent provides a reference point that may influence her choices*"

- $\Theta = \{(\Diamond, \Diamond), (\Diamond, c), (c, \Diamond), (c, c)\}$.
    - $\Diamond$ stands for not having a status-quo,
    - c stands for status-quo being coal

Consider the individual choices of Alice and Bob from (the subsets) of $X = \{c, n, s\}$ given below:

S	$C_A^{(\Diamond,\Diamond)}$	$C_B^{(\Diamond,\Diamond)}$	$C_A^{(\Diamond,c)}$	$C_B^{(\Diamond,c)}$	$C_A^{(c,\Diamond)}$	$C_B^{(c,\Diamond)}$	$C_A^{(c,c)}$	$C_B^{(c,c)}$
$\{c, n, s\}$	$\{s\}$	$\{s\}$	$\{s\}$	$\{n\}$	$\{n\}$	$\{s\}$	$\{n\}$	$\{n\}$
$\{c, n\}$	$\{n\}$	$\{n\}$	$\{n\}$	$\{n\}$	$\{n\}$	$\{n\}$	$\{n\}$	$\{n\}$
$\{c, s\}$	$\{s\}$	$\{s\}$	$\{s\}$	$\{c\}$	$\{c\}$	$\{s\}$	$\{c\}$	$\{c\}$
$\{n, s\}$	$\{s\}$	$\{s\}$	$\{s\}$	$\{s\}$	$\{s\}$	$\{s\}$	$\{n\}$	$\{n\}$

Application of the Corollary:

A social planner wants to ex-post implement the SCF f, **a particular selection from the BR-optimal outcomes**, described below:

State	$(\Diamond, \Diamond)$	$(\Diamond, c)$	$(c, \Diamond)$	(c, c)
BR-optimal	$\{s\}$	$\{n, s\}$	$\{n, s\}$	$\{n\}$
$\boldsymbol{f}$	s	s	s	n

The social planner **breaks the tie** in favor of s whenever n and s are **both** BR-optimal.

Application of the Corollary:

The SCF f satisfies **the consistency-unanimity property** since

- the collections and $\mathbb{S}_A := \{S_A(f, \Diamond), S_A(f, c)\}$ and $\mathbb{S}_B := \{S_B(f, \Diamond), S_B(f, c)\}$ such that
  - $S_A(f, \Diamond) = \{n, s\}$ and $S_A(f, c) = \{c, n, s\}$,
  - $S_B(f, \Diamond) = \{n, s\}$ and $S_B(f, c) = \{c, n, s\}$, and
- $\bar{X} = \{c, n, s\}$

satisfy conditions (i), (ii), and (iii) of the Corollary.

- That is, f is **ex-post implementable**.
  - The social planner can use the mechanism that we construct for our sufficiency results, if not a simpler one!

Simple Mechanisms

Why Simple? / What is Simple?

- There has been a recent interest in simple mechanisms in the literature e.g., Li (2017), Borgers and Li (2018).
- Dealing with individuals having limited cognitive abilities or behavioral biases **increases** the relevance and importance of the **simplicity** of mechanisms.
- We take the point of view that the smaller the number of messages in a mechanism the simpler the mechanism.

Back to the Motivating Example

Bob

$$\text{Alice}\qquad \begin{array}{c|ccc} & L & M & R \\ \hline U & n & c & n \\ M & c & s & c \\ D & n & s & s \end{array}$$

The following collections of sets satisfy the necessary conditions.
—as they are **induced** by the mechanism above.

$$\mathbb{S}_A = \{\{c, n\}, \{c, s\}, \{c, n, s\}\}$$
$$\mathbb{S}_B = \{\{c, n\}, \{c, s\}, \{n, s\}\}.$$

where

$$\text{Alice} \qquad \begin{array}{cc} S_A(f, \rho_B) = \{c, n\} & S_A(f, \gamma_B) = \{c, n, s\} \\ S_A(f', \rho_B) = \{c, s\} & S_A(f', \gamma_B) = \{c, s\} \end{array}$$

$$\text{Bob} \qquad \begin{array}{cc} S_B(f, \rho_A) = \{c, n\} & S_B(f, \gamma_A) = \{n, s\} \\ S_B(f', \rho_A) = \{c, s\} & S_B(f', \gamma_A) = \{c, n\}. \end{array}$$

Motivating Example: Revealed

Bob

Alice		$\{c,n\}$	$\{c,s\}$	$\{c,n,s\}$
	$\{c,n\}$	n	c	n
	$\{c,s\}$	c	s	c
	$\{n,s\}$	n	s	s

- Alice must have **at least** 3 strategies to be able to generate the opportunity set {c,n,s}.
- Bob must have **at least** 3 strategies to be able to let Alice generate all elements in $\mathbb{S}_A$. ($\#\mathbb{S}_A = 3$.)
- This is the "**simplest**" mechanism **given** the specific consistent collections $\mathbb{S}_A$ and $\mathbb{S}_B$.

Motivating Example: Can we go simpler?

Recall that two-individual consistency pinpoints 5 out of 8 sets!

- $S_A(f, \gamma_B) = \{c, n, s\}$; $S_A(f', \rho_B) = \{c, s\}$; $S_A(f', \gamma_B) = \{c, s\}$,
- $S_B(f, \gamma_A) = \{n, s\}$; $S_B(f, \rho_A) = \{c, n\}$.

The two-individidual consistent collections should be of the form:

- $\mathbb{S}_A = \{S_A(f, \rho_B), \{c, n, s\}, \{c, s\}\}$ and
- $\mathbb{S}_B = \{\{c, n\}, \{n, s\}, S_B(f', \rho_A), S_B(f', \gamma_A)\}$.

Motivating Example: Can we go simpler?

The two-individidual consistent collections should be of the form:

- $\mathbb{S}_A = \{S_A(f, \rho_B), \{c, n, s\}, \{c, s\}\}$ and
- $\mathbb{S}_B = \{\{c, n\}, \{n, s\}, S_B(f', \rho_A), S_B(f', \gamma_A)\}$.

Therefore,

- Alice must have **at least** 3 strategies in any mechanism that ex-post implements F —to be able to generate {c,n,s}!
- The **best** we can hope for Bob is to have 2 strategies —to be able to let Alice generate $\{c, n, s\}$ and $\{c, s\}$!

Motivating Example: Can we go simpler?

Let us try a 3×2 mechanism:

		Bob	
		$\{c, n, s\}$	$\{c, s\}$
Alice	$\{c, n\}$	x	c
	$\{n, s\}$	y	s
	$\{\mathrm{t}, \mathrm{z}\}$	z	t

- Since $\{x, y, z\} = \{c, n, s\}$, we must have $x \neq y \neq z$.
- But also, we must have $x = n$ and $y = n$ —for Bob to be able to generate $\{c, n\}$ and $\{n, s\}$ respectively. **Contradiction!**

Motivating Example: Can we go simpler?

No! We **cannot** go any simpler.

Corollary

Given the individual choices of Alice and Bob, **any** *mechanism that ex-post implements the SCS F must have* **at least three** *messages for Alice and the total number of messages for both players must be* **at least six***.*

- In this regard, there **does not exist** any simpler mechanism than the one below:

		Bob		
		L	*M*	*R*
Alice	*U*	*n*	*c*	*n*
	M	*c*	*s*	*c*
	D	*n*	*s*	*s*

Motivating Example: Simplest Mechanism is not unique!

The mechanism above is **not** the unique simplest mechanism!

- The following mechanism also works.

			Bob	
		$\{c, n\}$	$\{c, s\}$	$\{c, n, s\}$
Alice	$\{c\}$	c	c	c
	$\{c, n\}$	n	c	n
	$\{n, s\}$	n	s	s

The **only** difference is that $S_B(f', \rho_A) = \{c\}$ instead of $\{c, s\}$ now.

- The consistent collections $\mathbb{S}'_A = \{\{c, n\}, \{c, s\}, \{c, n, s\}\}$ $\mathbb{S}'_B = \{\{c\}, \{c, n\}, \{n, s\}\}$ also work!

Lower bounds on simplicity

- There could be **different** collections of sets consistent with the same SCS —as is the case for our motivating example.
  - How many sets there are in a collection, and how small these sets are, turn out to be **important** when designing simple mechanisms.
- Let $\{\mathbb{S}^\gamma\}_{\gamma\in\Gamma}$ be the set of all the collections of sets that satisfy consistency (or two-individual consistency).
  - $\mathbb{S}^\gamma = (\mathbb{S}_i^\gamma)_{i\in N}$ for $\gamma \in \Gamma$ where $\mathbb{S}_i^\gamma = \{S_i^\gamma(f, \theta_{-i}) | f \in F, \theta_{-i} \in \Theta_{-i}\}$.

Lower bounds on simplicity

(i) If $\mathbb{S}^\gamma$ is the **consistent** collection, i must have **at least** as many messages as the cardinality of the **maximal** set in $\mathbb{S}_i^\gamma$.

(ii) For each set in $\mathbb{S}_i^\gamma$, there must exist a particular message profile of **the individuals other than** i that let i to generate this set.

- If the collection $\mathbb{S}^\gamma$ happens to be the collection of opportunity sets generated by the mechanism that ex-post implements F, then
  - by (i) and (ii), the total number of messages must be **at least** $\max_{i\in N}[\#\mathbb{S}_i^\gamma + \max_{S\in\mathbb{S}_i^\gamma} \#S]$,
  - by (i), the total number of messages must be also **more than** $\sum_{i\in N} \max_{S\in\mathbb{S}_i^\gamma} \#S$.

(iii) Combining, the total number of messages must **exceed** $\max\left\{\min_{\gamma\in\Gamma} \max_{i\in N}[\#\mathbb{S}_i^\gamma + \max_{S\in\mathbb{S}_i^\gamma} \#S], \min_{\gamma\in\Gamma} \sum_{i\in N} \max_{S\in\mathbb{S}_i^\gamma} \#S\right\}$.

Lower bounds on simplicity

Theorem

In **any** *mechanism that* **ex-post implements** *the social choice set F,*

(i) *the minimum number of messages required for individual i is* $\boldsymbol{\min_{\gamma\in\Gamma}\max_{S\in\mathbb{S}_i^\gamma}\#S}$*,*

(ii) *the minimum number of message profiles required for the individuals other than i is* $\boldsymbol{\min_{\gamma\in\Gamma}\#\mathbb{S}_i^\gamma}$*, and*

(iii) *the minimum number of total messages required for all individuals is* $\max\left\{\min_{\gamma\in\Gamma}\max_{i\in N}[\#\mathbb{S}_i^\gamma+\max_{S\in\mathbb{S}_i^\gamma}\#S],\min_{\gamma\in\Gamma}\sum_{i\in N}\max_{S\in\mathbb{S}_i^\gamma}\#S\right\}$.

Thank you!