Implementation with a Sympathizer

Ozan Altuğ Altun[1] Mehmet Barlo[2] Nuh Aygün Dalkıran[3]

February 10, 2021

[1] Sabancı University
[2] Sabancı University
[3] Bilkent University

Our Main Question

Can a planner implement a given goal without knowing individuals' state-contingent preferences/choices?

That is, can we attain implementation of

- a goal that depends on states of the economy, while
- the planner is completely ignorant of how states of the economy and individuals' preferences are associated?

The Framework

- A planner aims to implement a goal that depends on *states of the economy*.
- Each state of the economy is associated with individuals' underlying preferences that we refer to as *payoff-relevant states*.
- The planner is completely ignorant of the association between the states of the economy and the payoff-relevant states.

The Framework

- A planner aims to implement a goal that depends on *states of the economy*.
- Each state of the economy is associated with individuals' underlying preferences that we refer to as *payoff-relevant states*.
- The planner is completely ignorant of the association between the states of the economy and the payoff-relevant states.

  For example, the planner could be an implementation consulting agency (e.g., McKinsey Implementation (McKinsey, 2018)) that is tasked to

  - elicit information about the details of a client firm and to implement a given policy contingent on this information.

The Framework

- A planner aims to implement a goal that depends on *states of the economy*.

- Each state of the economy is associated with individuals' underlying preferences that we refer to as *payoff-relevant states*.

- The planner is completely ignorant of the association between the states of the economy and the payoff-relevant states.

  Alternatively, the planner could be a court-appointed trustee authorized to run a company during its bankruptcy proceedings.

The Framework

- A planner aims to implement a goal that depends on *states of the economy*.
- Each state of the economy is associated with individuals' underlying preferences that we refer to as *payoff-relevant states*.
- The planner is completely ignorant of the association between the states of the economy and the payoff-relevant states.
- The planner can be viewed to face an extreme form of missing data.

What we do

In this framework, we investigate **Nash implementability** under **complete information**. We describe *situations* in which the planner (she) infers that she can

- **elicit** the relevant information about the association between the states of the economy and the payoff-relevant states from the individuals (he) unanimously, and
- use this information to **implement** the given collective goal.

Model

- X is the finite set of *alternatives*; $\mathcal{X}$ denotes the set of all *non-empty subsets* of X; N denotes the finite *society*.

To model the disparity in **knowledge**, we adopt the following:

- Ω is the set of *payoff-relevant states*;
- Θ is the set of the *states of the economy*;
- The *identification function* is a mapping $\pi^* : \Theta \to \Omega$;
- $R_i^\omega \subset X \times X$ denotes the *preferences* of agent $i \in N$ at $\omega \in \Omega$; $(R_i^\omega)_{i \in N, \omega \in \Omega}$ is in one-to-one correspondence with Ω. Moreover, $L_i^\omega(x) \equiv \{y \mid xR_i^\omega y\}$; $\mathbb{L}_i^\omega(x) \equiv \{S \in \mathcal{X} \mid S \subset L_i^\omega(x) \text{ and } x \in S\}$.

Information/Knowledge Assumptions

The **information and knowledge requirements** of our model are:

(*i*) the planner knows N, X, Ω, Θ, and $f : \Theta \to \mathcal{X}$; and

(*ii*) N, X, Ω, Θ, $\pi^* : \Theta \to \Omega$, $f : \Theta \to \mathcal{X}$, and the realized state of the economy $\theta \in \Theta$ are common knowledge among the individuals; and

(*iii*) items (*i*) and (*ii*) are common knowledge among the individuals and the planner.

The essence of the asymmetry of information between the planner and the individuals involves the identification function π^* and the realized state of the economy θ.

Nash Implementation

- The **social choice correspondence** (SCC) is given by $f : \Theta \to \mathcal{X}$.
- A **mechanism** is $\mu = (A, g)$ where
  - A_i is i's set of **actions** with $A \equiv \times_{i \in N} A_i$ and $A_{-i} \equiv \times_{j \neq i} A_j$,
  - $g : A \to X$ is the **outcome function**.
  - Given $a_{-i} \in A_{-i}$, i's **opportunity set** is $O_i^\mu(a_{-i}) \equiv g(A_i, a_{-i})$.
- $a^* \in A$ is a **Nash equilibrium (NE) of μ at $\omega \in \Omega$** if

$$g(a^*) \in \cap_{i \in N} C_i^\omega(O_i^\mu(a^*_{-i})),$$

  where for any $S \in \mathcal{X}$, $C_i^\omega(S) \equiv \{x \in S \mid x R_i^\omega y, \forall y \in S\}$.

  $NE^\mu(\theta)$ consists of $x \in X$ such that there is $a^* \in A$ with $g(a^*) = x$ and a^* is a Nash equilibrium of μ at $\pi^*(\theta) \in \Omega$.
- An SCC f is **Nash implementable by μ** if $\forall \theta \in \Theta$, $f(\theta) = NE^\mu(\theta)$.

Rational-Consistency: A Necessary Condition

Definition

A profile of sets $\mathbf{S} \equiv (S_i(x,\theta))_{i\in N,\ \theta\in\Theta,\ x\in f(\theta)}$ is **rational-consistent** with the given SCC $f : \Theta \to \mathcal{X}$ if

(i) $\forall i \in N$, $\forall \theta \in \Theta$, $\forall x \in f(\theta)$, $S_i(x,\theta) \in \mathbb{L}_i^{\pi^*(\theta)}(x)$; and

(ii) If $x \in f(\theta) \setminus f(\tilde{\theta})$, then $\exists j \in N$ s.t. $S_j(x,\theta) \notin \mathbb{L}_j^{\pi^*(\tilde{\theta})}(x)$.

$\mathcal{S}(f)$ denotes the set of profiles of sets that are rational-consistent with f.

Rational-consistency constitutes a variant of monotonicity of Maskin (1999) and the rational version of consistency of de Clippel (2014).

Rational-Consistency: A Necessary Condition

In words, a profile of sets $\mathbf{S}$ is **rational-consistent** with a given SCC f, if

(i) for every individual i and state of the economy θ and alternative x in $f(\theta)$, x is one of the best alternatives according to $R_i^{\pi^*(\theta)}$ in the set $S_i(x,\theta)$; and

(ii) if x is f-optimal at θ but not at $\tilde{\theta}$, then there exists $j \in N$ such that x is not among the best alternatives according to $R_j^{\pi^*(\tilde{\theta})}$ in $S_j(x,\theta)$.

A Necessity Theorem

Theorem

If the planner **knows** *that* **the SCC** $f : \Theta \to \mathcal{X}$ **is Nash implementable**, *then the planner* **infers** *that* $\mathcal{S}(f) \neq \emptyset$ **without necessarily knowing the full specification of sets that appear in** $\mathcal{S}(f)$**.**

Significance of the theorem:

- This result is a reevaluation of the necessity theorems of
  - Maskin (1999) and
  - de Clippel (2014)

  related to the knowledge inferred by a completely ignorant planner.

A Necessity Theorem

Theorem

If the planner **knows** *that* **the SCC $f : \Theta \to \mathcal{X}$ is Nash implementable**, *then the planner* **infers** *that* $\mathcal{S}(f) \neq \emptyset$ **without necessarily knowing the full specification of sets that appear in $\mathcal{S}(f)$.**

Sketch of the Proof:

- Suppose the planner knows that μ^* implements SCC f. Then, she infers that $\forall\theta$ and $\forall x \in f(\theta)$ there is an NE a^x s.t. $g(a^x) = x$.

A Necessity Theorem

Theorem

If the planner **knows** *that* **the SCC $f : \Theta \to \mathcal{X}$ is Nash implementable**, *then the planner* **infers** *that* $\mathcal{S}(f) \neq \emptyset$ **without necessarily knowing the full specification of sets that appear in $\mathcal{S}(f)$.**

Sketch of the Proof:

- Suppose the planner knows that μ^* implements SCC f. Then, she infers that $\forall \theta$ and $\forall x \in f(\theta)$ there is an NE a^x s.t. $g(a^x) = x$.
- The planner does not know $\pi^*(\theta)$ and may be uncertain about a^x if μ^* delivers x via at least two action profiles. Still, she infers that

$$x \in \cap_{i \in N} C_i^{\pi^*(\theta)}(O_i^{\mu^*}(a_{-i}^x)). \tag{1}$$

A Necessity Theorem

Theorem

If the planner **knows** *that* **the SCC $f : \Theta \to \mathcal{X}$ is Nash implementable**, *then the planner* **infers** *that* **$\mathcal{S}(f) \neq \emptyset$ without necessarily knowing the full specification of sets that appear in $\mathcal{S}(f)$.**

Sketch of the Proof:

- Suppose the planner knows that μ^* implements SCC f. Then, she infers that $\forall\theta$ and $\forall x \in f(\theta)$ there is an NE a^x s.t. $g(a^x) = x$.
- The planner does not know $\pi^*(\theta)$ and may be uncertain about a^x if μ^* delivers x via at least two action profiles. Still, she infers that

$$x \in \cap_{i \in N} C_i^{\pi^*(\theta)}(O_i^{\mu^*}(a^x_{-i})). \tag{1}$$

- The planner considers inferences about $\mathbf{S}^{\mu^*} \equiv (S_i^{\mu^*}(x,\theta))_{i,\theta,x\in f(\theta)}$, where $S_i^{\mu^*}(x,\theta) = O_i^{\mu^*}(a^x_{-i})$, without knowing its full specification.

A Necessity Theorem

Theorem

If the planner **knows** *that* **the SCC $f : \Theta \to \mathcal{X}$ is Nash implementable**, *then the planner* **infers** *that* $\mathcal{S}(f) \neq \emptyset$ **without necessarily knowing the full specification of sets that appear in $\mathcal{S}(f)$.**

Sketch of the Proof:

- Then, thanks to (1) the planner infers that $\mathbf{S}^{\mu^*}$ is s.t.

$$x \in \cap_{i \in N} C_i^{\pi^*(\theta)}(S_i^{\mu^*}(x,\theta)) \Rightarrow S_i^{\mu^*}(x,\theta) \in \mathbb{L}_i^{\pi^*(\theta)}(x), \forall i \in N.$$

- So, the planner deduces that $\mathbf{S}^{\mu^*}$ satisfies (i) of rational-consistency even though she does not know its full specification.

A Necessity Theorem

Theorem

If the planner **knows** *that* **the SCC $f : \Theta \to \mathcal{X}$ is Nash implementable**, *then the planner* **infers** *that* $\mathcal{S}(f) \neq \emptyset$ **without necessarily knowing the full specification of sets that appear in $\mathcal{S}(f)$.**

<u>Sketch of the Proof</u>:

- If the planner knows that $x \in f(\theta) \setminus f(\tilde{\theta})$, she infers that
$$S_i^{\mu^*}(x, \theta) \in L_i^{\pi^*(\tilde{\theta})}(x), \forall i \in N \tag{2}$$
would imply an impasse:
- The planner contemplating on $\mathbf{S}^{\mu^*}$, without knowing its full specification, infers that (2) implies a^x is an NE at $\pi^*(\tilde{\theta})$.
- As μ^* Nash implements f, she deduces that x would be in $f(\tilde{\theta})$; a contradiction.

A Necessity Theorem

Theorem

If the planner **knows** *that* **the SCC $f : \Theta \to \mathcal{X}$ is Nash implementable***, then the planner* **infers** *that* $\mathcal{S}(f) \neq \emptyset$ **without necessarily knowing the full specification of sets that appear in $\mathcal{S}(f)$.**

Implications:

- The planner knowing the existence of a profile rational-consistent with the given SCC $f : \Theta \to \Omega$, constitutes the minimal information
  - about the association between Θ and Ω
  - in conjunction with the Nash implementability of f.
- That is why the planner's knowledge of $\mathcal{S}(f) \neq \emptyset$ can be regarded as a *feasibility* requirement for the Nash implementability of f.
- So, sufficiency has to involve the planner's knowledge of $\mathcal{S}(f) \neq \emptyset$.

A Necessity Theorem

Theorem

If the planner **knows** *that* **the SCC $f : \Theta \to \mathcal{X}$ is Nash implementable**, *then the planner* **infers** *that* $\mathcal{S}(f) \neq \emptyset$ **without necessarily knowing the full specification of sets that appear in $\mathcal{S}(f)$.**

Implications:

- Even if the planner were to know $\mathcal{S}(f) \neq \emptyset$, she does not know π^*.
- She needs to know the full specification of a rational-consistent profile $\mathbf{S} \in \mathcal{S}(f)$ to design mechanisms that can implement f in NE.
- We consider mechanisms in which the planner *asks* agents' help.
- We show that she can elicit a rational-consistent profile $\mathbf{S}$ from the society unanimously with the help of a *sympathizer*.

Sympathy

The notion of **sympathy** is defined for

- mechanisms involving the announcements of a profile of choice sets
- by modifying *partial honesty* of Dutta and Sen (2012) so that
- it involves only announcements of profiles of sets and not states.

Guidance Mechanisms

To define the notion of **sympathy** for any given SCC $f : \Theta \to \Omega$ consider

- **guidance mechanisms** $\mathcal{M}^S$ consisting of $\mu^S = (A^S, g^S)$ where $A_i^S = \mathcal{S} \times M_i^S$ for all $i \in N$ where $\mathcal{S}$ is given by $\{\mathbf{S} = (S_i(x,\theta))_{i,\theta,x \in f(\theta)} \mid \forall i \in N, \forall \theta \in \Theta, \forall x \in f(\theta), x \in S_i(x,\theta)\}$, and M_i^S is a message set of $i \in N$.
- We note that $\mathcal{S}(f) \subset \mathcal{S}$.
- An **action** of individual i in a guidance mechanism $\mu^S \in \mathcal{M}^S$ is
$$a_i = (\mathbf{S}, m_i) \in \mathcal{S} \times M_i, \quad \text{and}$$
consists of a **profile** of sets feasible w.r.t. f and some **messages**.

Sympathy

For any $f : \Theta \to \mathcal{X}$, any $\mu \in \mathcal{M}^S$, and any $\omega \in \Omega$, the correspondence $BR_i^\omega : A_{-i} \twoheadrightarrow A_i$ identifies i's **best responses** at ω to others' actions.

- If i is a **sympathizer** of f at $\omega \in \Omega$, then $\forall a_{-i} \in A_{-i}$

  (i) $\mathbf{S} \in \mathcal{S}(f)$, $\tilde{\mathbf{S}} \notin \mathcal{S}(f)$, and $m_i \in M_i$ implies $(\mathbf{S}, m_i) \in BR_i^\omega(a_{-i})$ and $(\tilde{\mathbf{S}}, m_i) \notin BR_i^\omega(a_{-i})$ if

  $$g((\mathbf{S}, m_i), a_{-i}) \; R_i^\omega \; g(a_i', a_{-i}) \; \forall a_i' \in A_i, \text{ and}$$
  $$g((\tilde{\mathbf{S}}, m_i), a_{-i}) \; R_i^\omega \; g(a_i'', a_{-i}) \; \forall a_i'' \in A_i; \text{ and}$$

  (ii) in all other cases,

  $$a_i \in BR_i^\omega(a_{-i}) \text{ if and only if } g(a_i, a_{-i}) \; R_i^\omega \; g(a_i', a_{-i}) \; \forall a_i' \in A_i.$$

- If i is **not a sympathizer** of f at $\omega \in \Omega$, then $\forall a_{-i} \in A_{-i}$,

  $$a_i \in BR_i^\omega(a_{-i}) \text{ if and only if } g(a_i, a_{-i}) \; R_i^\omega \; g(a_i', a_{-i}) \; \forall a_i' \in A_i.$$

Sympathy

Guidance mechanisms involve some messages besides the announcement of a profile of sets for each individual.

A **sympathizer** of the SCC at a payoff-relevant state is an individual who

- strictly prefers the action that consists of the announcement of a consistent profile of sets coupled with some messages
- to another action which involves the announcement of an inconsistent profile while he continues to send the same messages
- whenever both actions provide this agent's top alternative in his opportunity set given others' actions.
- In all other cases, he is a regular economic agent who chooses as described by his state-contingent preferences.

The Sympathizer Property

- We say that the environment satisfies the **sympathizer property** with respect to SCC f
- if for all $\omega \in \Omega$, there is at least one sympathizer of f at ω,
- while the identity of each sympathizer of f at ω is privately known only by himself.

Nash* Equilibrium: NE with Sympathy

- Given $\mu \in \mathcal{M}$, $a^* \in A$ is a **Nash* equilibrium** of μ at ω if for all $i \in N$, $a_i^* \in BR_i^{\omega}(a_{-i}^*)$, where the best responses are as above.

- Given SCC f, if $\mu \in \mathcal{M} \setminus \mathcal{M}^S$ or there are no sympathizers of f at ω, then NE of μ at ω and Nash* equilibrium of μ at ω coincide.

- An SCC $f : \Theta \to \mathcal{X}$ is **implementable by a mechanism $\mu \in \mathcal{M}$ in Nash* equilibrium**, if
  - (i) for any $\theta \in \Theta$ and $x \in f(\theta)$, there exists $a^x \in A$ such that $g(a^x) = x$ and $a_i^x \in BR_i^{\pi^*(\theta)}(a_{-i}^x)$ for all $i \in N$; and
  - (ii) for any $\theta \in \Theta$, $a^* \in A$ with $a_i^* \in BR_i^{\pi^*(\theta)}(a_{-i}^*)$ for all $i \in N$ implies $g(a^*) \in f(\theta)$.

Economic Environments

The **economic environment** assumption demands that there is some weak form of disagreement in the society at every payoff-relevant state:

For every alternative and for every state,

- there are at least two individuals who do not choose that alternative at that state from the set of all alternatives.

This assumption is also used in Bergemann and Morris (2008), Kartik and Tercieux (2012), Barlo and Dalkıran (2020).

Our Main Result

Theorem

Suppose $n \geq 3$. Suppose that the planner knows that

(*i*) *the* **environment is economic**, *and the* **sympathizer property** *holds, and*

(*ii*) *the SCC $f : \Theta \to \mathcal{X}$ has a* **rational-consistent** *profile of sets, i.e., $\mathcal{S}(f) \neq \emptyset$, while she does not necessarily know the full specification of the sets that appear in $\mathcal{S}(f)$.*

Then, the planner infers that f is **Nash* implementable by a guidance mechanism** $\mu \in \mathcal{M}^S$, *and for any $\theta \in \Theta$ and any Nash* equilibrium $\bar{a} = (\bar{\mathbf{S}}^{(i)}, \bar{m}_i)_{i \in N}$ of μ at $\pi^*(\theta)$,*

$$\bar{\mathbf{S}}^{(i)} = \mathbf{S} \text{ for some } \mathbf{S} \in \mathcal{S}(f) \text{ for all } i \in N.$$

Our Main Result

If the planner (she) knows that there are least three individuals and

(*i*) the **environment is economic**, and the **sympathizer property** holds, and

(*ii*) the SCC f has a **rational-consistent** profile of sets (the full specification of which she does not know),

then she infers that f is

- **Nash implementable** by a mechanism that
- **elicits the relevant information**

about the association between the payoff-relevant states and the states of the economy from the society unanimously, while

- the identity of the sympathizer is known only to himself.

Our Main Result: The Mechanism

- i's **action** is $a_i = (\mathbf{S}^{(i)}, \theta^{(i)}, x^{(i)}, k^{(i)}) \in A_i$ with $\mathbf{S}^{(i)} \in \mathcal{S}$, $\theta^{(i)} \in \Theta$, $x^{(i)} \in X$, and $k^{(i)} \in \mathbb{N}$; let $m_i = (\theta^{(i)}, x^{(i)}, k^{(i)})$.

- The **outcome function** is defined via the rules specified as follows:

Rule 1 : $g(a) = x$ — if $\mathbf{S}^{(i)} = \mathbf{S}$ for all $i \in N \setminus \{i'\}$ for some $i' \in N$, and $m_j = (\theta, x, \cdot)$ for all $j \in N$ with $x \in f(\theta)$,

Rule 2 : $g(a) = \begin{cases} x' & \text{if } x' \in S_j(x,\theta) \\ & \text{where } S_j(x,\theta) = \mathbf{S}|_{j,\theta,x\in f(\theta)}, \\ x & \text{otherwise.} \end{cases}$ — if $\mathbf{S}^{(i)} = \mathbf{S}$ for all $i \in N \setminus \{i'\}$ for some $i' \in N$, and $m_i = (\theta, x, \cdot)$ for all $i \in N \setminus \{j\}$ with $x \in f(\theta)$, and $m_j = (\theta', x', \cdot) \neq (\theta, x, \cdot)$,

Rule 3 : $g(a) = x^{(i^*)}$ where $i^* = \min\{j \in N \mid k^{(j)} = \max_{i' \in N} k^{(i')}\}$ — otherwise.

Our Main Result: Sketch of the Proof

The arguments behind the proof:

- The **planner** does **not know** the identification function $\pi^* : \Theta \to \Omega$ and hence **agents' lower contour sets.**

- She **uses** agents' **announcements** to construct their **opportunity sets** whenever announcements of all individuals but one coincide.

  Agents' **opportunity sets** are determined by their announcements

  $$(\mathbf{S}^{(i)})_{i \in N} \in \mathcal{S}^N.$$

- The **sympathizer property** ensures that all agents announce a rational-consistent profile in any Nash* equilibrium under Rule 1.

- The **economic environment** assumption ensures that there is no Nash* equilibrium under Rules 2 and 3.

Inference of Rational-Consistency

- We propose a way to ensure the planner's inference of the existence of a rational-consistent profile. It involves Maskin monotonicity.

- A correspondence $\phi : \Omega \to 2^X$ is **Maskin monotonic** if $x \in \phi(\omega)$ and $L_i^{\omega}(x) \subseteq L_i^{\tilde{\omega}}(x)$ for all $i \in N$ implies $x \in \phi(\tilde{\omega})$, where $\omega, \tilde{\omega} \in \Omega$
  - 2^X denotes the set of all (possibly empty) subsets of X.

- $f_\Omega : \Omega \to 2^X$ is an **extension of an SCC $f : \Theta \to \mathcal{X}$ to Ω** if $f(\theta) = f_\Omega(\pi^*(\theta))$ for all $\theta \in \Theta$.

Inference of Rational-Consistency

Proposition

If the planner knows that SCC $f : \Theta \to \mathcal{X}$

- *has a* **Maskin monotonic extension** *even if she does not know the full specification of this extension,*

then she infers that $\mathcal{S}(f) \neq \emptyset$ **without** *necessarily* **knowing the specification of sets** *that appear in* $\mathcal{S}(f)$.

Sympathy versus Honesty

- Now, we adopt the convention that
  - a state under complete information is to encompass all the information that is common knowledge among the individuals.
- The set of **grand states**: $\Sigma \equiv \{(\theta, \omega, \pi) \in \Theta \times \Omega \times \Pi \mid \pi(\theta) = \omega\}$, with $\sigma = (\theta, \omega, \pi)$ with $\pi(\theta) = \omega$ and $\Pi \equiv \{\pi' \mid \pi' : \Theta \to \Omega\}$.
  - We extend f onto Σ by $f(\sigma) = f(\theta)$ for all $\sigma = (\theta, \omega, \pi) \in \Sigma$.
- The mechanisms with the announcement of a grand state, $\mathcal{M}^{\Sigma}$:
  - $\mu = (A, g)$ with $A_i = (\sigma^{(i)}, m_i) \in \Sigma \times M_i$, $\forall i \in N$.

Partial Honesty

Individuals' **best responses** at $\sigma = (\theta, \omega, \pi)$ under partial honesty are:

- If i is **partially honest** at $\sigma = (\theta, \omega, \pi)$, then $\forall a_{-i} \in A_{-i}$

  (i) $\tilde{\sigma} \neq \sigma$ and $m_i, \tilde{m}_i \in M_i$ implies $(\sigma, m_i) \in BR_i^{\omega}(a_{-i})$ and $(\tilde{\sigma}, \tilde{m}_i) \notin BR_i^{\omega}(a_{-i})$ if

  $$g((\sigma, m_i), a_{-i}) \; R_i^{\omega} \; g(a_i', a_{-i}) \; \forall a_i' \in A_i, \text{ and}$$
  $$g((\tilde{\sigma}, \tilde{m}_i), a_{-i}) \; R_i^{\omega} \; g(a_i'', a_{-i}) \; \forall a_i'' \in A_i; \text{ and}$$

  (ii) in all other cases,

  $$a_i \in BR_i^{\omega}(a_{-i}) \text{ if and only if } g(a_i, a_{-i}) \, R_i^{\omega} \, g(a_i', a_{-i}) \, \forall a_i' \in A_i.$$

- If i is **not partially honest** at σ, then $\forall a_{-i} \in A_{-i}$,

  $$a_i \in BR_i^{\omega}(a_{-i}) \text{ if and only if } g(a_i, a_{-i}) \; R_i^{\omega} \; g(a_i', a_{-i}) \; \forall a_i' \in A_i.$$

Partial Honesty (in words)

$\mathcal{M}^{\Sigma}$ consists of mechanisms that involve some messages besides the announcement of a grand state for each individual.

A **partially honest individual** at the realized grand state

- strictly prefers the action that consists of the announcement of the realized grand state coupled with some messages
- to another action which involves the announcement of a different grand state while sending some other messages
- whenever both actions provide this agent's top alternative in his opportunity set given others' actions.
- In all other cases, he is a regular economic agent who chooses as described by his state-contingent preferences.

Partial Honesty: An Implication

Theorem 1 of Dutta and Sen (2012) implies:

If the planner knows that for every state $\sigma \in \Sigma$

- there is a partially honest individual at σ (even if she does not know the identity of this agent) for all $\sigma \in \Sigma$, and that
- the SCC $f : \Theta \to \mathcal{X}$ satisfies the no-veto property,

then she infers that

- f is Nash* implementable and in every such equilibrium all but one announce a grand set *aligned* with the realized grand set.

Weak Partial Honesty

Next, we define **weak partial honesty** to compare with sympathy s.t. the agent at hand is partially honest with respect to π but not (θ, ω):

Agents' **best responses** at $\sigma = (\theta, \omega, \pi^*)$ under weak partial honesty are

- If i is **weakly partially honest** at $\sigma = (\theta, \omega, \pi^*)$, then $\forall a_{-i} \in A_{-i}$

  (i) $\tilde{\sigma} = (\tilde{\theta}, \tilde{\omega}, \pi^*)$ and $\hat{\sigma} = (\hat{\theta}, \hat{\omega}, \hat{\pi})$ with $\hat{\pi} \neq \pi^*$ and $\tilde{m}_i, \hat{m}_i \in M_i$ implies $(\tilde{\sigma}, \tilde{m}_i) \in BR_i^{\omega}(a_{-i})$ and $(\hat{\sigma}, \hat{m}_i) \notin BR_i^{\omega}(a_{-i})$ if

  $$g((\tilde{\sigma}, \tilde{m}_i), a_{-i}) \; R_i^{\omega} \; g(a_i', a_{-i}) \; \forall a_i' \in A_i, \text{ and}$$
  $$g((\hat{\sigma}, \hat{m}_i), a_{-i}) \; R_i^{\omega} \; g(a_i'', a_{-i}) \; \forall a_i'' \in A_i; \text{ and}$$

  (ii) in all other cases,

  $$a_i \in BR_i^{\omega}(a_{-i}) \text{ if and only if } g(a_i, a_{-i}) \, R_i^{\omega} \, g(a_i', a_{-i}) \, \forall a_i' \in A_i.$$

- If i is **not partially honest** at σ, then $\forall a_{-i} \in A_{-i}$,

  $$a_i \in BR_i^{\omega}(a_{-i}) \text{ if and only if } g(a_i, a_{-i}) \; R_i^{\omega} \; g(a_i', a_{-i}) \; \forall a_i' \in A_i.$$

Weak Partial Honesty (in words)

A **weakly partially honest agent** at the realized grand state, where the realized (true) association between the states of the economy and payoff-relevant states is π^* as in our setup,

- strictly prefers the action that consists of the announcement of the realized association, π^*, coupled with some messages
- to another action which involves the announcement of a different association, $\pi \neq \pi^*$, while sending some other messages
- whenever both actions provide this agent's top alternative in his opportunity set given others' actions.
- In all other cases, he is a regular economic agent who chooses as described by his state-contingent preferences.

Weak Partial Honesty vs. Strong Sympathy

If the realized association between the states of the economy Θ and payoff-relevant states Ω is π^* as in our setup, then

- the definition of a weakly partially honest individual resembles our definition of a **strong sympathizer**.

Given a guidance mechanism $\mu^S = (A^S, g^S)$ the **best responses** of a **strong sympathizer** i of f at ω are such that if $\forall a_{-i}$,

(i) $\mathbf{S} \in \mathcal{S}(f)$, $\hat{\mathbf{S}} \notin \mathcal{S}(f)$, and $m_i, \hat{m}_i \in M_i^S$ implies $(\mathbf{S}, m_i) \in BR_i^{\omega}(a_{-i})$ and $(\hat{\mathbf{S}}, \hat{m}_i) \notin BR_i^{\omega}(a_{-i})$ if

- $g^S((\mathbf{S}, m_i), a_{-i}) R_i^{\omega} g^S(a_i', a_{-i})$ for all $a_i' \in A_i$, and
- $g^S((\hat{\mathbf{S}}, \hat{m}_i), a_{-i}) R_i^{\omega} g^S(a_i'', a_{-i})$ for all $a_i'' \in A_i$; and

(ii) otherwise, $a_i \in BR_i^{\omega}(a_{-i})$ iff $g^S(a_i, a_{-i}) R_i^{\omega} g^S(a_i', a_{-i})$ $\forall a_i' \in A_i$.

If $i \in N$ is **not a strong sympathizer** of f at ω, then $a_i \in BR_i^{\omega}(a_{-i})$ iff $g^S(a_i, a_{-i}) R_i^{\omega} g^S(a_i', a_{-i})$ for all $a_i' \in A_i$.

Strong Sympathy (in words)

Guidance mechanisms involve some messages besides the announcement of a profile of sets for each individual. A **strong sympathizer** of the SCC at a payoff-relevant state is an individual who

- strictly prefers the action that consists of the announcement of a consistent profile of sets coupled with some messages
- to another action which involves the announcement of an inconsistent profile along with some other messages
- whenever both actions provide this agent's top alternative in his opportunity set given others' actions.
- In all other cases, he is a regular economic agent who chooses as described by his state-contingent preferences.

We note that every **strong sympathizer** is a **sympathizer**.

Weak Partial Honesty vs. Strong Sympathy

If the realized association between the states of the economy Θ and payoff-relevant states Ω is π^* as in our setup, then

- the definition of a weakly partially honest individual resembles our definition of a **strong sympathizer**.

Observations:

- Interchanging π^* with $\mathbf{S} \in \mathcal{S}(f)$ and $\hat{\pi}$ with $\hat{\mathbf{S}} \notin \mathcal{S}(f)$ displays parallels between strong sympathy of f at $\pi^*(\theta)$ and weak partial honesty at $\sigma = (\theta, \omega, \pi^*)$.

- If the planner is informed of π^*, then she can construct the set of rational-consistent profiles $\mathcal{S}(f)$.

- But, she cannot necessarily identify π^* uniquely if she is informed of an element $\mathbf{S}$ in $\mathcal{S}(f)$.

Weak Partial Honesty vs. Strong Sympathy

If the realized association between the states of the economy Θ and payoff-relevant states Ω is π^* as in our setup, then

- the definition of a weakly partially honest individual resembles our definition of a **strong sympathizer**.

Observations:

- Thus, in our construct, to implement a given SCC,

  the **extent of information** the planner seeks to elicit with the **help** of a **strong sympathizer** is

  less than

  the **extent of information** the planner obtains **thanks** to a **weakly partially honest individual**.

Behavioral Formulation

We extend our setting and results to the **behavioral domain**:

- The **choice** of i at $\omega \in \Omega$ is $C_i^\omega : \mathcal{X} \to \mathcal{X}$ s.t. $C_i^\omega(S) \subset S, \forall S \in \mathcal{X}$.
- A profile $\mathbf{S} := (S_i(x,\theta))_{i,\theta,\ x \in f(\theta)}$ is **consistent** with the given SCC $f : \Theta \to \mathcal{X}$ if

  (*i*) $\forall \theta \in \Theta, \forall x \in f(\theta)$, $x \in \cap_{i \in N} C_i^{\pi^*(\theta)}(S_i(x,\theta))$; and

  (*ii*) $x \in f(\theta) \setminus f(\theta')$ implies $x \notin \cap_{i \in N} C_i^{\pi^*(\theta')}(S_i(x,\theta))$.

  $\mathcal{S}(f)$ denotes the set of all profiles of sets that are consistent with f.

Using de Clippel's necessity result and similar arguments leading to our necessity theorem, enable us to conclude the following:

- If the planner knows that f is behavioral Nash implementable, then she infers that $\mathcal{S}(f) \neq \emptyset$ without necessarily knowing the full specification of sets that appear in $\mathcal{S}(f)$.

Behavioral Sympathy

For any $f : \Theta \to \mathcal{X}$, any $\mu \in \mathcal{M}^S$, and any $\omega \in \Omega$, the correspondence $BR_i^\omega : A_{-i} \twoheadrightarrow A_i$ identifies i's **behavioral best responses** at ω to a_{-i}.

- If i is a **behavioral sympathizer** of f at ω, then $\forall a_{-i}$

  (i) $\mathbf{S} \in \mathcal{S}(f)$, $\tilde{\mathbf{S}} \notin \mathcal{S}(f)$, and $m_i \in M_i$ implies $(\mathbf{S}, m_i) \in BR_i^\omega(a_{-i})$ and $(\tilde{\mathbf{S}}, m_i) \notin BR_i^\omega(a_{-i})$ if

  $$g((\mathbf{S}, m_i), a_{-i}), g((\tilde{\mathbf{S}}, m_i), a_{-i}) \in C_i^\omega(O_i^\mu(a_{-i}); \text{ and}$$

  (ii) in all other cases,

  $$a_i \in BR_i^\omega(a_{-i}) \text{ if and only if } g(a_i, a_{-i}) \in C_i^\omega(O_i^\mu(a_{-i})).$$

- If i is **not a behavioral sympathizer** of f at ω, then $\forall a_{-i}$

  $$a_i \in BR_i^\omega(a_{-i}) \text{ if and only if } g(a_i, a_{-i}) \in C_i^\omega(O_i^\mu(a_{-i})).$$

Strong Behavioral Sympathy

For any $f : \Theta \to \mathcal{X}$, any $\mu \in \mathcal{M}^S$, and any $\omega \in \Omega$, the correspondence $BR_i^\omega : A_{-i} \twoheadrightarrow A_i$ identifies i's **behavioral best responses** at ω to a_{-i}.

- If i is a **strong behavioral sympathizer** of f at ω, then $\forall a_{-i}$

  (i) $\mathbf{S} \in \mathcal{S}(f)$, $\tilde{\mathbf{S}} \notin \mathcal{S}(f)$, and $m_i, \tilde{m}_i \in M_i$ implies $(\mathbf{S}, m_i) \in BR_i^\omega(a_{-i})$ and $(\tilde{\mathbf{S}}, \tilde{m}_i) \notin BR_i^\omega(a_{-i})$ if

  $$g((\mathbf{S}, m_i), a_{-i}), g((\tilde{\mathbf{S}}, \tilde{m}_i), a_{-i}) \in C_i^\omega(O_i^\mu(a_{-i}); \text{ and}$$

  (ii) in all other cases,

  $$a_i \in BR_i^\omega(a_{-i}) \text{ if and only if } g(a_i, a_{-i}) \in C_i^\omega(O_i^\mu(a_{-i})).$$

- If i is **not a strong behavioral sympathizer** of f at ω, then $\forall a_{-i}$

  $$a_i \in BR_i^\omega(a_{-i}) \text{ if and only if } g(a_i, a_{-i}) \in C_i^\omega(O_i^\mu(a_{-i})).$$

Behavioral Sympathizer Property

The environment satisfies the **behavioral sympathizer property** with respect to SCC f if for all $\omega \in \Omega$,

- there is at least **one behavioral sympathizer** of f at ω,
- while the identity of each behavioral sympathizer of f at ω is privately known only by himself.

Behavioral Sympathizer Property

The environment satisfies the **behavioral sympathizer property** with respect to SCC f if for all $\omega \in \Omega$,

- there is at least **one behavioral sympathizer** of f at ω,
- while the identity of each behavioral sympathizer of f at ω is privately known only by himself.

The environment satisfies the **strong behavioral sympathizer property** with respect to SCC f if for all $\omega \in \Omega$,

- there is are least **two strong behavioral sympathizers** of f at ω,
- while the identity of each strong behavioral sympathizer of f at ω is privately known only by himself.

Societal Agreement Conditions

We say that

(i) the environment features **societal non-satiation** if for all ω and all x, there is an individual i s.t. $x \notin C_i^{\omega}(X)$.

(ii) the **behavioral economic environment** assumption holds if for all ω and all x, there are i, j with $i \neq j$ s.t. $x \notin C_i^{\omega}(X) \cup C_j^{\omega}(X)$.

(iii) an SCC $f : \Theta \to \mathcal{X}$ satisfies the **behavioral no-veto property** if for all θ, $x \in \cap_{i \in N \setminus \{j\}} C_i^{\pi^*(\theta)}(X)$ for some $j \in N$ implies $x \in f(\theta)$.

We remark that

- Behavioral economic environment implies societal non-satiation.
- Every SCC satisfies the behavioral no-veto property vacuously whenever the behavioral economic environment assumption holds.
- When the meaning is clear, we **do not spell out the 'behavioral' label** of our notions and in our results.

Sufficiency in Noneconomic Environments

Theorem

Let $n \geq 3$ and the SCC $f : \Theta \to \mathcal{X}$ be given. Suppose that

(*i*) *the planner knows that* **societal non-satiation** *and the* **strong sympathizer property** *hold, and*

(*ii*) *without necessarily knowing the full specification of sets that appear in $\mathcal{S}(f)$, the planner knows that*

(*ii*.1) *$\mathcal{S}(f) \neq \emptyset$, i.e. f has a* **consistent** *profile of sets and that*

(*ii*.2) *f satisfies the* **no-veto property**.

Then, the planner infers that f **is Nash* implementable by a guidance mechanism** *$\mu \in \mathcal{M}^S$, and for any state of the economy $\theta \in \Theta$ and any Nash* equilibrium $\bar{a} = (\bar{\mathbf{S}}^{(i)}, \bar{m}_i)_{i \in N}$ of μ at state $\pi^*(\theta)$,*

$$\bar{\mathbf{S}}^{(i)} = \mathbf{S} \text{ for some } \mathbf{S} \in \mathcal{S}(f) \text{ for all } i \in N \setminus \{j\} \text{ for some } j \in N.$$

Sufficiency in Noneconomic Environments

If the planner (she) knows that there are least three individuals and

(*i*) the **societal non-satiation**, and the **strong sympathizer property** holds, and

(*ii*) f has a **rational-consistent** profile of sets (the full specification of which she does not know) and satisfies the **no-veto property**,

then she infers that f is

- **Nash implementable** by a guidance mechanism that
- **elicits the information about consistency**

from the society *almost* unanimously, while

- the identity of the sympathizer is known only to himself.

The Mechanism

- We use the same **mechanism**:

  i's **action** is $a_i = (\mathbf{S}^{(i)}, \theta^{(i)}, x^{(i)}, k^{(i)}) \in A_i$ with $\mathbf{S}^{(i)} \in \mathcal{S}$, $\theta^{(i)} \in \Theta$, $x^{(i)} \in X$, and $k^{(i)} \in \mathbb{N}$; let $m_i = (\theta^{(i)}, x^{(i)}, k^{(i)})$.

  The **outcome function** is defined via the rules specified as follows:

Rule 1 : $g(a) = x$ if $\mathbf{S}^{(i)} = \mathbf{S}$ for all $i \in N \setminus \{i'\}$ for some $i' \in N$, and $m_j = (\theta, x, \cdot)$ for all $j \in N$ with $x \in f(\theta)$,

Rule 2 : $g(a) = \begin{cases} x' & \text{if } x' \in S_j(x,\theta) \\ & \text{where } S_j(x,\theta) = \mathbf{S}|_{j,\theta,x\in f(\theta)}, \\ x & \text{otherwise.} \end{cases}$ if $\mathbf{S}^{(i)} = \mathbf{S}$ for all $i \in N \setminus \{i'\}$ for some $i' \in N$, and $m_i = (\theta, x, \cdot)$ for all $i \in N \setminus \{j\}$ with $x \in f(\theta)$, and $m_j = (\theta', x', \cdot) \neq (\theta, x, \cdot)$,

Rule 3 : $g(a) = x^{(i^*)}$ where $i^* = \min\{j \in N \mid k^{(j)} = \max_{i' \in N} k^{(i')}\}$ otherwise.

The Sketch of the Proof

The arguments behind the proof:

- The planner **uses** agents' **announcements** to construct their **opportunity sets** if announcements of all but one coincide.

  Agents' **opportunity sets** are determined by their announcements

  $$(\mathbf{S}^{(i)})_{i \in N} \in \mathcal{S}^N.$$

- The dismissal of the **economic environment** and the adoption of **societal non-satiation** coupled with the **no-veto property** result in new Nash* equilibria under Rules 1 and 2.

- The **sympathizer property** does not suffice to ensure the extraction of information about consistency unanimously.

- Now, we need at least two strong sympathizers, and hence the **strong sympathizer property**, to elicit the information about consistency **almost** unanimously.

The Sketch of the Proof

The arguments behind the proof:

- To see why we need **two strong sympathizers**, consider the following case under Rule 1 which we have to rule out in Nash* eq.:
- All i except the odd man out j announce the same profile $\mathbf{S} \notin \mathcal{S}(f)$ while the remaining messages of all (including j) coincide and contain θ and x such that x is f-optimal at θ.
- Then, we cannot rule this case out under Rule 1 in Nash* eq. with only one sympathizer as he could be the odd man out. So, we need another sympathizer announcing $\mathbf{S}$. But that is not enough because:
- This agent has to be a strong sympathizer to obtain a profitable deviation by changing his profile announcement and integer choice. Thus, we also need to strengthen sympathy to strong sympathy.
- Even then there are Nash* equilibria in which not all, but all except one, announce a consistent profile of sets.

Concluding Remarks

- We investigate **Nash implementation** when the planner is **completely ignorant** of the association between the states of the economy and individuals' preferences.

- We identify **conditions** ensuring that the planner infers that she can
  - **elicit** the relevant information about the association between the states of the economy and the payoff-relevant states from the society unanimously, and
  - use this information to **implement** the given collective goal.