ELICITING AWARENESS

Evan Piermont
Royal Holloway, University of London

Aarhus Workshop on Known and Unknown Unknowns

July 2023

- A decision maker must choose a ‘plan-of-action;’ what action to take provided the future resolution of uncertainty
- He is unaware of some relevant contingencies and *knows this is possible*
- He can seek the council of an expert who is more aware than himself

Why is the interesting?

- When preferences are not aligned, the expert might strategically conceal her awareness
- Can the dm do anything to incentivize revelation?
- Importantly, even with full/complete contracting, the dm cannot articulate what he wants
- A(n unaware) designer may not be able to solve the problem, if mechanisms depend on the unknowns

Example

- A politician (the decision maker) is trying to write environmental legislation that
  - can be contingent on the future realized environmental state-of-affairs, but
  - can depend only on those contingencies he is aware of.
- He can enlist the help of an environmental scientist (the expert) who may reveal what she is aware of

Example

- The true state-space is $\Omega = \{\omega, \nu\}$; equally likely
- Set of actions $\mathcal{A} = \{a, b, c\}$
- The politician must choose legislation $\mathfrak{c} : \Omega \to \mathcal{A}$

Example

The expert can tell distinguish the states, but the politician cannot.

$$\mathscr{P}_e = \{\{\omega\}, \{\nu\}\} \qquad \mathscr{P}_d = \{\{\omega, \nu\}\}$$

$$u_d = \left\{ \begin{array}{c|c|c|c} & a & b & c \\ \hline \omega & 4 & 0 & 2 \\ \hline \nu & 0 & 6 & 0 \end{array} \right.$$

$$u_e = \left\{ \begin{array}{c|c|c|c} & a & b & c \\ \hline \omega & 0 & 2 & 4 \\ \hline \nu & 0 & 2 & 4 \end{array} \right.$$

$$u_d = \left\{ \begin{array}{c|c|c|c} & a & b & c \\ \hline \omega & 4 & 0 & 2 \\ \hline \nu & 0 & 6 & 0 \end{array} \right.$$

$$u_d = \left\{ \begin{array}{c|c|c|c} & a & b & c \\ \hline \{\omega, \nu\} & 2 & 3 & 1 \end{array} \right.$$

$$u_e = \left\{ \begin{array}{c|c|c|c} & a & b & c \\ \hline \omega & 0 & 2 & 4 \\ \hline \nu & 0 & 2 & 4 \end{array} \right.$$

$$u_e = \left\{ \begin{array}{c|c|c|c} & a & b & c \\ \hline \{\omega, \nu\} & 0 & 2 & 4 \end{array} \right.$$

$$u_d = \left\{ \begin{array}{c|c|c|c} & a & b & c \\ \hline \omega & 4 & 0 & 2 \\ \hline \nu & 0 & 6 & 0 \end{array} \right.$$

$$u_e = \left\{ \begin{array}{c|c|c|c} & a & b & c \\ \hline \omega & 0 & 2 & 4 \\ \hline \nu & 0 & 2 & 4 \end{array} \right.$$

What would the politician implement:

$$u_d = \left\{ \begin{array}{c|c|c|c} & a & b & c \\ \hline \omega & 4 & 0 & 2 \\ \hline \nu & 0 & 6 & 0 \end{array} \right.$$

$$u_e = \left\{ \begin{array}{c|c|c|c} & a & b & c \\ \hline \omega & 0 & 2 & 4 \\ \hline \nu & 0 & 2 & 4 \end{array} \right.$$

What would the politician implement:

- Without revelation $\mathfrak{c} = b$
- $\mathbb{E}[u_d] = 3, \mathbb{E}[u_e] = 2$

$$u_d = \left\{ \begin{array}{c|c|c|c} & a & b & c \\ \hline \omega & 4 & 0 & 2 \\ \hline \nu & 0 & 6 & 0 \end{array} \right.$$

$$u_e = \left\{ \begin{array}{c|c|c|c} & a & b & c \\ \hline \omega & 0 & 2 & 4 \\ \hline \nu & 0 & 2 & 4 \end{array} \right.$$

What would the politician implement:

- With revelation: $\mathfrak{c}' : \begin{cases} \omega \mapsto a \\ \nu \mapsto b \end{cases}$
- $\mathbb{E}[u_d] = 5, \mathbb{E}[u_e] = 1$; So the expert won't reveal.

$$u_d = \left\{ \begin{array}{c|c|c|c} & a & b & c \\ \hline \omega & 4 & 0 & 2 \\ \hline \nu & 0 & 6 & 0 \end{array} \right. \qquad u_e = \left\{ \begin{array}{c|c|c|c} & a & b & c \\ \hline \omega & 0 & 2 & 4 \\ \hline \nu & 0 & 2 & 4 \end{array} \right.$$

- But, $\mathfrak{c}^\star : \begin{cases} \omega \mapsto c \\ \nu \mapsto b \end{cases}$ is a Pareto improvement over no revelation
- $\mathbb{E}[u_d] = 4, \mathbb{E}[u_e] = 3$

Example

- The Pareto improvement $\mathfrak{c}^{\star}$, requires revelation
- But revealing allows the politician to exploit the expert
- What is the politician could commit:
  - Propose $\mathfrak{c} = b$ (his outside option)
  - After the expert reveals, propose some other contract $\mathfrak{c}^{\dagger}$
  - $\mathfrak{c}^{\dagger}$ only get implemented if the expert agrees

Example

Internalizing this, the politician solves:

$$\max_{\mathfrak{c}^\dagger:\Omega\to\mathcal{A}} \mathbb{E}[u_d(\mathfrak{c}^\dagger)] \qquad \text{subject to} \qquad \mathbb{E}[u_e(\mathfrak{c}^\dagger)] \geq \mathbb{E}[u_e(\mathfrak{c})] \qquad \text{(IC)}$$

Example

Internalizing this, the politician solves:

$$\max_{\mathfrak{c}^\dagger : \Omega \to \mathcal{A}} \mathbb{E}[u_d(\mathfrak{c}^\dagger)] \qquad \text{subject to} \qquad \mathbb{E}[u_e(\mathfrak{c}^\dagger)] \geq \mathbb{E}[u_e(\mathfrak{c})] \qquad \text{(IC)}$$

- The solution is $\mathfrak{c}^\star : \begin{cases} \omega \mapsto c \\ \nu \mapsto b \end{cases}$

So a two stage game with commitment to never revoke prior proposals resulted in

- full revelation
- an efficient contract

So a two stage game with commitment to never revoke prior proposals resulted in

- full revelation
- an efficient contract

Does this always work?

So a two stage game with commitment to never revoke prior proposals resulted in

- full revelation
- an efficient contract

Does this always work? No

Example

- Take $\Omega = \{\omega, \nu, \upsilon\}$; equally likely
- Set of actions $\mathcal{A} = \{a, b, c, d\}$
- The expert can tell distinguish the states, but the decision maker cannot.

$$\mathscr{P}_e = \{\{\omega\}, \{\nu\}, \{\upsilon\}\} \qquad \mathscr{P}_d = \{\{\omega, \nu, \upsilon\}\}$$

$$u_d = \left\{ \begin{array}{c|c|c|c|c} & a & b & c & d \\ \hline \omega & 2 & 0 & 0 & 0 \\ \hline \nu & 2 & 3 & 0 & 0 \\ \hline \upsilon & 2 & 2 & 4 & 3 \end{array} \right.$$

$$u_e = \left\{ \begin{array}{c|c|c|c|c} & a & b & c & d \\ \hline \omega & 1 & 2 & 0 & 0 \\ \hline \nu & 1 & 2 & 1 & 0 \\ \hline \upsilon & 1 & 2 & 1 & 3 \end{array} \right.$$

$$u_d = \left\{ \begin{array}{c|c|c|c|c} & a & b & c & d \\ \hline \omega & 2 & 0 & 0 & 0 \\ \hline \nu & 2 & 3 & 0 & 0 \\ \hline \upsilon & 2 & 2 & 4 & 3 \end{array} \right.$$

$$u_e = \left\{ \begin{array}{c|c|c|c|c} & a & b & c & d \\ \hline \omega & 1 & 2 & 0 & 0 \\ \hline \nu & 1 & 2 & 1 & 0 \\ \hline \upsilon & 1 & 2 & 1 & 3 \end{array} \right.$$

- Without revelation $\mathfrak{c} = a$
- $\mathbb{E}[u_d] = \frac{6}{3}, \mathbb{E}[u_e] = \frac{3}{3}$

$$u_d = \left\{ \begin{array}{c|c|c|c|c} & a & b & c & d \\ \hline \omega & 2 & 0 & 0 & 0 \\ \hline \nu & 2 & 3 & 0 & 0 \\ \hline \upsilon & 2 & 2 & 4 & 3 \end{array} \right.$$

$$u_e = \left\{ \begin{array}{c|c|c|c|c} & a & b & c & d \\ \hline \omega & 1 & 2 & 0 & 0 \\ \hline \nu & 1 & 2 & 1 & 0 \\ \hline \upsilon & 1 & 2 & 1 & 3 \end{array} \right.$$

- With full revelation: $\mathfrak{c}' : \begin{cases} \omega \mapsto a \\ \nu \mapsto b \\ \upsilon \mapsto c \end{cases}$
- $\mathbb{E}[u_d] = \frac{9}{3}, \mathbb{E}[u_e] = \frac{4}{3}$

$$u_d = \left\{ \begin{array}{c|c|c|c|c} & a & b & c & d \\ \hline \omega & 2 & 0 & 0 & 0 \\ \hline \nu & 2 & 3 & 0 & 0 \\ \hline \upsilon & 2 & 2 & 4 & 3 \end{array} \right.$$

$$u_e = \left\{ \begin{array}{c|c|c|c|c} & a & b & c & d \\ \hline \omega & 1 & 2 & 0 & 0 \\ \hline \nu & 1 & 2 & 1 & 0 \\ \hline \upsilon & 1 & 2 & 1 & 3 \end{array} \right.$$

- Revealing $\{\{\omega\}, \{\nu, \upsilon\}\}$: $\mathfrak{c}''$: $\begin{cases} \omega & \mapsto a \\ \{\nu, \upsilon\} & \mapsto b \end{cases}$
- $\mathbb{E}[u_d] = \frac{7}{3}, \mathbb{E}[u_e] = \frac{5}{3}$
- So, even with commitment, the expert does not fully reveal

$$u_d = \left\{ \begin{array}{c|c|c|c|c} & a & b & c & d \\ \hline \omega & 2 & 0 & 0 & 0 \\ \hline \nu & 2 & 3 & 0 & 0 \\ \hline \upsilon & 2 & 2 & 4 & 3 \end{array} \right.$$

$$u_e = \left\{ \begin{array}{c|c|c|c|c} & a & b & c & d \\ \hline \omega & 1 & 2 & 0 & 0 \\ \hline \nu & 1 & 2 & 1 & 0 \\ \hline \upsilon & 1 & 2 & 1 & 3 \end{array} \right.$$

- Again, this is inefficient: $\mathfrak{c}^\star : \begin{cases} \omega \mapsto a \\ \nu \mapsto b \\ \upsilon \mapsto d \end{cases}$
- $\mathbb{E}[u_d] = \frac{8}{3}, \mathbb{E}[u_e] = \frac{6}{3}$

What if the procedure was repeated?

(1) Initially, the decision maker commits to $\mathfrak{c} = a$

(1) Initially, the decision maker commits to $\mathfrak{c} = a$

(2) The expert partially reveals $\big\{\{\omega\}, \{\nu, \upsilon\}\big\}$

(1) Initially, the decision maker commits to $\mathfrak{c} = a$

(2) The expert partially reveals $\big\{\{\omega\}, \{\nu, \upsilon\}\big\}$

(3) The decision maker proposes to $\mathfrak{c}''$; the expert accepts

(1) Initially, the decision maker commits to $\mathfrak{c} = a$

(2) The expert partially reveals $\{\{\omega\}, \{\nu, \upsilon\}\}$

(3) The decision maker proposes to $\mathfrak{c}''$; the expert accepts

- This solves the IC-constrained maximization problem with constraint $\mathfrak{c}$

(1) Initially, the decision maker commits to $\mathfrak{c} = a$

(2) The expert partially reveals $\{\{\omega\}, \{\nu, \upsilon\}\}$

(3) The decision maker proposes to $\mathfrak{c}''$; the expert accepts

- This solves the IC-constrained maximization problem with constraint $\mathfrak{c}$
- This becomes the new IC constraint

(1) Initially, the decision maker commits to $\mathfrak{c} = a$

(2) The expert partially reveals $\{\{\omega\}, \{\nu, \upsilon\}\}$

(3) The decision maker proposes to $\mathfrak{c}''$; the expert accepts

- This solves the IC-constrained maximization problem with constraint $\mathfrak{c}$
- This becomes the new IC constraint

(4) The expert reveals the rest of her awareness

(1) Initially, the decision maker commits to $\mathfrak{c} = a$

(2) The expert partially reveals $\big\{\{\omega\}, \{\nu, \upsilon\}\big\}$

(3) The decision maker proposes to $\mathfrak{c}''$; the expert accepts

- This solves the IC-constrained maximization problem with constraint $\mathfrak{c}$
- This becomes the new IC constraint

(4) The expert reveals the rest of her awareness

(5) The decision maker solves

$$\max_{\mathfrak{c}^\dagger : \Omega \to \mathcal{A}} \mathbb{E}[u_d(\mathfrak{c}^\dagger)] \qquad \text{subject to} \qquad \mathbb{E}[u_e(\mathfrak{c}^\dagger)] \geq \mathbb{E}[u_e(\mathfrak{c}'')] \qquad \text{(IC)}$$

(1) Initially, the decision maker commits to $\mathfrak{c} = a$

(2) The expert partially reveals $\{\{\omega\}, \{\nu, \upsilon\}\}$

(3) The decision maker proposes to $\mathfrak{c}''$; the expert accepts

- This solves the IC-constrained maximization problem with constraint $\mathfrak{c}$
- This becomes the new IC constraint

(4) The expert reveals the rest of her awareness

(5) The decision maker solves

$$\max_{\mathfrak{c}^\dagger : \Omega \to \mathcal{A}} \mathbb{E}[u_d(\mathfrak{c}^\dagger)] \qquad \text{subject to} \qquad \mathbb{E}[u_e(\mathfrak{c}^\dagger)] \geq \mathbb{E}[u_e(\mathfrak{c}'')] \tag{IC}$$

- $\mathfrak{c}^\star$ is the solution

Does this always work?

Does this always work? Yes

Does this always work? Yes

- This iterated procedure (ending when nothing novel is revealed) always results in:
    - full revelation

Does this always work? Yes

- This iterated procedure (ending when nothing novel is revealed) always results in:
  - full revelation
  - an efficient contract

Does this always work? Yes

- This iterated procedure (ending when nothing novel is revealed) always results in:
  - full revelation
  - an efficient contract
- More remarkably: any procedure yielding full revelation is equivalent to this one

The latter point can be seen as an impossibility result:

- Without commitment to leave proposed contracts on the table, full revelation cannot be obtained.

Each proposed contract specifies:

(1) The outcome should the game end
  - dm wants to maximize his own payoff

(2) The incentive constraint should the game continue
  - dm wants to minimize the expert's payoff

In the examples, contracts solved (1) ignoring (2)

- The dm cannot conceive of what the expert is aware of
- It seems prudent, therefore, to consider *robust* strategies:
  - A strategy is *robust* when it maximizes the worst case outcome (over what set?)
- Robust strategies turn out to be exactly those that follow the principle of myopic optimization

Model

How does the dm reason about the expert?

Hypothetical State-Space

Call $h = (W, (v_d, v_e), p)$ a **hypothetical states-space**,

- W is a finite set
- $v_i : W \times \mathcal{A} \to \mathbb{R}$, for $i \in \{d, e\}$, and,
- $p \in \Delta(W)$

Let $\mathscr{H}$ collect all hss; $\mathscr{H}$ are the possible types of expert

Say that $h' = (W', (v'_d, v'_e), p')$ **refines** $h = (W, (v_d, v_e), p)$, written $h \preccurlyeq h'$, if there exists a surjection $q : W' \to W$ such that

$$\sum_{w' \in q^{-1}(w)} p'(w') = p(w) \qquad \text{for all } w \in W \tag{H1}$$

and

$$\sum_{w' \in q^{-1}(s)} w'_i(w', \cdot) p'(w') = v_i(w, \cdot) \qquad \text{for all } w \in W \text{ and } i \in \{d, e\} \tag{H2}$$

Strategies

Let h_0 be the awareness of the dm:

- The actions of the expert (of type h) are all sequences of revelations $H = (h_0, h_1, \ldots h_n)$ where $h_i \prec h_{i+1} \preccurlyeq h$.
- A strategy of the dm is a mapping from sequences of revelations to contracts

  $$\sigma : (h_0, h_1, \ldots h_n) \mapsto \mathfrak{c}$$

  where $\mathfrak{c} : W_n \to \mathcal{A}$.

Strategies

Let h_0 the type of the dm and h the type of the expert with $h = (W, (v_d, v_e), p)$.

Best Response

Call $H = (h_0, h_1, \ldots, h_n)$ a best response to strategy σ if

$$\mathbb{E}(v_e(\sigma(H)) \geq \mathbb{E}(v_e(\sigma(H'))$$

for any other feasible sequence of revelations, H'.

Full Revelation

Theorem

The following are equivalent:

(1) σ is equivalent[1] to some $\bar{\sigma}$ that globally satisfies the IC constraint:

$$\mathbb{E}(v_e(\bar{\sigma}(H)) \geq \mathbb{E}(v_e(\bar{\sigma}(H')) \qquad \text{whenever } H \text{ extends } H'. \tag{1}$$

(2) σ is **fully revealing**: for every type h, it is a best response to reveal h

[1] I.e., for all types the strategies yield the same set of best responses / revelations and same implemented contract

Robustness

The dm cannot properly envision hypothetical state spaces:

- It seems unreasonable to assume a probability over them
- Instead, the dm could maximize the *worst case outcome*
- Follows the literature on robustness mechanism design

Robustness

Robustness

Call an strategy, σ, **robust** if it is IC and for all (reachable) H

$$\inf_{\substack{H' \text{ extends } H \\ H' \text{best response to } \sigma}} \mathbb{E}(v_d(\sigma(H'))) \geq \inf_{H' \text{ extends } H} \mathbb{E}(v_d(\sigma'(H')))$$

for other σ' that coincides with σ up-to H.

Robustness

Theorem

The following are equivalent:

(1) σ is robust

(2) σ is equivalent to some $\bar{\sigma}$ that is myopically optimal: $\bar{\sigma}(H)$ maximizes $\mathbb{E}(v_d)$ subject to the IC constraint.

The Designers Problem

More generally, often awareness is decentralized:

- A designer wants the decision maker to take some action
- The designer does not know the dm's or the expert's awareness
- A **mechanism** elicits awareness and returns an action recommendation

Mechanism

A **mechanism** is a mapping from pairs of types into contracts:

$$\mathcal{M} : (h^d, h^e) \mapsto \mathcal{M}(h^d, h^e)$$

where $\mathcal{M}(h^d, h^e) : W(h^d \wedge h^e) \to \mathcal{A}$

Desiderata:

INDIVIDUAL RATIONALITY: the dm can not do better alone (there is no constraint for the expert)

INCENTIVE COMPATIBILITY: i prefers to report h^i than any $h \prec h^i$

PARETO OPTIMALITY: there is no feasible contract that dominates the outcome of the mechanism

These are all **ex-post** restrictions — they must hold for all type realizations

Consider the mechanism, $\mathcal{M}^*$, that enforces round-by-round commitment then implements the game described above.

Theorem

The mechanism $\mathcal{M}^*$

- is individually rational, incentive compatible, and Pareto optimal, and,
- $\mathbb{E}(v_d)$-dominates any other such mechanism (point-wise over the type-space)

- there is a ‘dual’ mechanism that is expert-optimal:
- it reverses the proposer and acceptor roles.
- requires only one round

Thank You!