# Belief elicitation

Elias Tsakas

Maastricht University

EGT course
July 2022

# Roadmap

# Background

- Theory has to be testable.
- To test EGT we need to be able to measure beliefs.
- However, beliefs are by definition latent.
- So the best we can hope for is to elicit them indirectly *using some incentive compatible mechanism*

# Elicitation mechanisms

- All mechanisms are essentially direct mechanisms:
    - Input: the agent's self-reported probability about the state space (i.e., the opponent's strategy space)
    - Output: an act (i.e., state-contingent lottery)
- Mechanisms differ mainly in how they are framed.
- (Strict) incentive compatibility: it is (uniquely) optimal to report truthfully.

# Roadmap

1. Background
2. Mechanisms
3. Additional issues

# Preliminaries

- Two-player strategic-form game: $I = \{a, b\}$ and $(S_i, u_i)_{i \in I}$
- Player $i$ has a (first order) belief $\mu_i \in \Delta(S_j)$.
- Of course beliefs are latent.
- Player $i$ self-reports some $r_i \in \Delta(S_j)$, not necessarily the same as $\mu_i$.

# Scoring rules

- A scoring rule is a function $\pi_i : \Delta(S_j) \to \mathbb{R}^{S_j}$
- The idea is that for each report $r_i$, player $i$ receives a payment $\pi_i(r_i, s_j)$ in case the strategy $s_j$ is actually played by player $j$.
- For the time being, assume that player $i$ is risk neutral, i.e., $u_i(x) = x$.
- Thus, $i$'s utility becomes $u_i(r_i, s_j) := u_i\big(\pi_i(r_i)\big)$
- So, if the report $r_i$ is submitted, player $i$'s expected utility is

$$\mathbb{E}_{\mu_i}\big(u_i(r_i, \cdot)\big) = \sum_{s_j \in S_j} \mu_i(s_j) \cdot u_i(r_i, s_j)$$

- The scoring rule is called proper, if expected utility is maximized when true belief is reported, i.e.,

$$\mathbb{E}_{\mu_i}\big(u_i(\mu_i, \cdot)\big) \geq \mathbb{E}_{\mu_i}\big(u_i(r_i, \cdot)\big)$$

- It is strictly proper if previous inequality is strict.

# Properness

- Savage provided a very useful characterization of properness.
- Every (strictly) proper scoring rule is characterized by a (strictly) convex function.
- Figure on the board
- Quadratic scoring rule.
- Discrete scoring rule: guess the opponent's strategy.
- Is non-incentivized elicitation a proper scoring rule?
- What do you understand by the following: the "more convex" a function is the stronger the incentives the scoring rule provides.

# Risk attitudes

- Strictly proper scoring rules are not incentive-compatible in general.
- Suppose that player $i$ is risk-loving, e.g., let $u_i(x) = x^2$
- Then, QSR does not guarantee truth-telling. Why?

# Binarized scoring rules

- Common solution is to linearize utilities by paying people in probability units to win a fixed prize.
- In thsi case, formally the scoring rule becomes a function $\pi_i : \Delta(S_j) \rightarrow [0,1]^{S_j}$
- The idea is that for each report $r_i$, player $i$ receives a good payoff $\overline{x}$ with probability $\pi_i(r_i, s_j)$ and a bad payoff $\underline{x}$ with the remaining probability.
- Normalize vNM utilities so that $u_i(\overline{x}) = 1$ and $u_i(\underline{x}) = 0$.
- Then, expected utility is equal to expected winning probability of $\overline{x}$
- So, if the report $r_i$ is submitted, player $i$'s expected winning probability is

$$\mathbb{E}_{\mu_i}(\pi_i(r_i, \cdot)) = \sum_{s_j \in S_j} \mu_i(s_j) \cdot \pi_i(r_i, s_j)$$

# Binarized scoring rules

- We can then define (strict) properness regardless of risk-attitudes.
- The binarized scoring rule is called proper, if expected winning probability is maximized when true belief is reported, i.e.,

$$\mathbb{E}_{\mu_i}\big(\pi_i(\mu_i,\cdot)\big) \geq \mathbb{E}_{\mu_i}\big(\pi_i(r_i,\cdot)\big)$$

- It is strictly proper if previous inequality is strict.
- Empirically the main problem with binarized scoring rules is that they involve compound acts (i.e., subjects need to combine two probability measures) which seems to be a difficult task for many people.

# Stochastic mechanisms

- Binarized scoring rules belong to a broader family of stochastic elicitation mechanisms.
- Stochastic elicitation mechanisms are direct mechanisms that pay each report with an AA act.
- Different mechanisms differ in how the payments are framed.]
- Stochastic mechanisms are incentive compatible even under weaker assumptions on the preferences (i.e., SEU is not needed; probabilistic sophistication suffices).

# Karni mechanism

- Take a strategy $s_j \in S_j$, and let $r_i(s_j)$ be $i$ reported belief that $s_j$ will be played by $j$.
- Then, we draw a random number $p \in [0, 1]$ (from uniform distribution).
  - If $r_i(s_j) \geq p$, then $i$ will receive $\overline{x}$ if $s_j$ is played and $\underline{x}$ otherwise
  - If $r_i(s_j) < p$, then $i$ will receive a lottery that pays $\overline{x}$ with probability $p$ and $\underline{x}$ with probability $1 - p$.
- This is essentially a BDM (it resembles a second price auction).
- There are dynamic versions which resemble clock auctions.

# Matching probabilities

- For each $p \in [0, 1]$, the player chooses between the following two options:
    - Receive $\overline{x}$ if $s_j$ is played and $\underline{x}$ otherwise
    - Receive a lottery that pays $\overline{x}$ with probability $p$ and $\underline{x}$ with probability $1 - p$.
- Finally, one of the choice problems will be randomly picked (i.e., one $p \in [0, 1]$) and the choice will be implemented.

# Roadmap

# Hedging

- Incentivizing the elicitation task means that player $i$ is paid for two different tasks that are related to each other.
- Take a pure coordination game with strategies $\{H, L\}$.
- Suppose that $i$ is risk-averse (i.e., $u_i(0) = 0$, $u_i(1) = 1$ and $u_i(2) = 1.1$), and faces a QBSR (with prize equal to 1 Euro).
- Suppose that $\mu_i(H) = \mu_i(L) = 1/2$.
- By incentive-compatibility of the BSR, she must report a uniform belief, which will pay her 1 Euro with probability 3/4 regardless of what the opponent does.
- Then, by playing $H$ she will get the lottery $(0.375 \times 2, 0.5 \times 1, 0.125 \times 0)$ which gives SEU of 0.91. The same is true by playing $L$.
- However, if she played $H$ and reported that the opponent will choose $L$ with probability 1, then she guarantees a payoff of 1, and utility 1.
- Solution: pay randomly one of the two tasks.