Ec717a: Robust Mechanism Design (Bergemann-Morris Ecta 2005)

Dilip Mookherjee

Boston University

Ec 717a, 2020: Lecture 2

Introduction: Problems with CPA

- Most game theory and mechanism design is based on the assumption of a common prior
- Based on Harsanyi ‘story’ that types are drawn from an urn according to a known probability distribution
- In words, CPA means ‘All differences in beliefs stem from differences in information’
- CPA has strong implications: we cannot ‘agree to disagree’ (Aumann); no-trade theorems (Milgrom and Stokey)

Payoff Types and CPA

- Standard mechanism design with private values — identifies types with 'valuations' (v_i) or payoff parameters, which agents observe privately
- Agent strategies $b_i(v_i)$; payoff functions $\Pi_i(b_i, b_{-i}; v_i)$
- With multiple agents that interact in the mechanism, agent's rational behavior defined by maximization of expected payoffs, with expectation taken over actions of other agents
- Optimal strategy of agent i therefore depends on:
  - i's beliefs $F_i(v_{-i})$ over others types
  - beliefs over strategies of others $b_j(v_j)$
- Beliefs over strategies of others, are rationalized by conjecture that others behave optimally —- requires i to know beliefs others (j) hold ($F_j(v_{-j})$)

CPA, contd

- Equilibrium strategies have to be common knowledge, and this requires players' beliefs to be common knowledge
- Departing from CPA: means we have to modify the definition of Bayesian equilibrium itself
- Have to allow players to have beliefs about beliefs of others which could be wrong
- This will significantly complicate the theory... is CPA a problem, or a convenient fiction?

Related Problems: Mechanism Design with Correlated Values

- Cremer-McLean (Ecta 1988) showed that in auctions where bidders have correlated values (hence beliefs depend on own type: $F_i(v_{-i}|v_i)$), it is (generically) possible to design an auction in which the auctioneer can extract almost all of every bidder's rents and thereby get arbitrarily close to the first-best (with dominant strategy equilibrium)
- But if valuations are independent, such rent extraction is not possible!
- McAfee-Reny (Ecta 1992) showed a similar result extends to most agency problems

Intuition for Cremer-McLean Result

- With correlated values, a change in v_i is associated with a change in bidder i's beliefs over v_{-i}, and this mapping is locally invertible
- It is possible (at the participation) for P to design a mechanism to elicit bidder 1's beliefs regarding bidder 2's value (ask bidder 1 to forecast what 2 will report, compare the forecast with the actual reports, construct side payments conditioned on these)
- P can invert from 1's beliefs to infer 1's valuation, and thus overcome the problem of private information
- With private (independent) values, this inversion is not possible, because bidder 1's beliefs do not vary with her own valuation

Ways of Escaping the Cremer-McLean Result

- Some people have pointed out that the result requires risk-neutrality, unlimited liability, absence of collusion among agents etc.
- Neeman (JET 2004) argued for another foundational reason built into the model: agents' types are defined by their values/preferences, and beliefs are a function of this type
- If we enlarge the notion of type to include both preferences and beliefs as separate components, beliefs would not determine values
- Two types could differ in values and have the same beliefs; then knowing beliefs will not reveal the agents' preferences, and agents can continue to earn private information rents that cannot be extracted even with correlated values

Bergemann-Morris (2005)

- BM argue for the need to extend mechanism design theory to contexts where we enlarge type spaces
- 'Types' should be multidimensional: (values, beliefs), full support priors should be allowed (echoing Neeman (2004))
- Moreover, require mechanisms to be robust to the possibility of misspecified priors (i.e., allow agents to hold 'wrong' beliefs)

Game Theory without common knowledge beliefs

- If we allow players to not know the beliefs of others, and continue with a Bayesian approach, have to model beliefs over beliefs of others
- Harsanyi's **universal type space**: a type includes valuations, beliefs over valuations of others, beliefs over beliefs of others (second-order beliefs), beliefs over second order beliefs (third order), ad infinitum
- The universal type space is (tautologically) common knowledge, given the description of the game; formalized by Mertens-Zamir (1985)
- $t_i \in T_i$: set of universal types of i; t_i includes i's valuation v_i, first order and all higher order beliefs

Payoff Environment

- Finite set of agents $i = 1, \ldots, I$
- Payoff type of i is $\theta_i \in \Theta_i$, a finite set; θ denotes $(\theta_1, \ldots, \theta_I)$
- Set of outcomes Y
- Utility function of i: $u_i : Y \times \Theta \to \Re$
- Planner's normative goal: in state θ, select any outcome in $F(\theta) \subseteq Y$, the *social choice correspondence*

Example: Separable Payoff Environment

- $Y = Y_0 \times Y_1 \times Y_2 \times ... Y_I$
- $u_i = v_i(y_0, y_i; \theta)$
- There is a function $f_0 : \Theta \to Y_0$ and $F_i(\theta) \subseteq Y_i$ such that

$$F(\theta) = f_0(\theta) \times F_1(\theta) \times \dots F_I(\theta)$$

- *Interpretation:* Player i cares only about public decision y_0 and a private i-specific outcome y_i; private outcomes for different agents can be chosen independently by P (rules out balanced budget constraint)

Example of a Separable Environment: Quasi-linear Case

- $Y = Y_0 \times Y_1 \times Y_2 \times \ldots Y_I$ where $Y_i = \Re$, set of possible transfers to i, and Y_0 is a set of possible 'allocations' of goods

- $U_i = v_i(y_0, \theta) + y_i$

- P cares only about the allocation, represented by a *social choice function (SCF)* $f_0 : \Theta \to Y_0$, so

$$F(\theta) = \{(y_0, y_1, \ldots, y_n) | y_0 = f_0(\theta)\}$$

- (No budget constraint restricting private transfers; could incorporate participation constraints if needed)

The Universal Type Space

- $t_i^0 = \theta_i(t_i) \in T_i^0$, i's payoff parameter
- $t_i^1 = (\theta_i, \pi_i^1(t_{-i}^0)) \in T_i^1$, where $\pi_i^1(t_{-i}^0)$ denotes i's beliefs over $t_j^0 \equiv \theta_j$, all $j \neq i$
- Proceed iteratively: $t_i^k = (\theta_i, \pi_i^k(t_{-i}^{k-1})) \in T_i^k$, all k
- Have to impose consistency/coherence conditions, across different orders
- Very large type space! If there is *some* common knowledge then the support of the beliefs could be restricted to a subspace

The Universal Type Space, contd.

- Under some conditions, the type space can be represented either by the infinite sequence of higher order beliefs, or implicitly as follows
- Type space $\mathcal{T} \equiv (T_i, \hat{\theta}_i, \hat{\pi}_i)_i$ where agent i's:
  - type is $t_i \in T_i$
  - payoff is $\hat{\theta}_i(t_i) : T_i \to \Theta_i$
  - belief is $\hat{\pi}_i(t_i) : T_i \to \Delta(T_{-i})$

Incentive Compatibility

- Strategies and beliefs are now a function of the (possibly infinite) type t_i
- Define expected payoffs for any type t_i of i relative to these beliefs: $E_{\hat{\pi}_i(t_i)[t_{-i}]} U_i(b, b_{-i}(t_{-i}); \hat{\theta}_i(t_i), \hat{\theta}_{-i}(t_{-i}))$, and extend the definition of a Bayesian equilibrium ($b_i(t_i)$ which maximizes this)
- The question of existence of a mechanism with an equilibrium that implements the desired SCC, can be simplified as usual by the Revelation Principle

(Interim) Incentive Compatibility

- Revelation Principle: wlog can focus on direct mechanisms $f : T \to Y$, specifying outcomes corresponding to any type report
- $f : T \to Y$ is *interim incentive compatible (IIC)* on type space T if $\forall i$:

$$\begin{aligned} & E_{\hat{\pi}_i(t_i)[t_{-i}]} U_i(f(t_i, t_{-i}), \hat{\theta}_i(t_i), \hat{\theta}_{-i}(t_{-i})) \\ \geq \quad & E_{\hat{\pi}_i(t_i)[t_{-i}]} U_i(f(t_i^{'}, t_{-i}), \hat{\theta}_i(t_i), \hat{\theta}_{-i}(t_{-i})) \end{aligned}$$

for all $t_i, t_i^{'} \in T_i$

- Larger the type space T_i, the more incentive constraints have to be satisfied

(Interim) Incentive Compatibility, contd.

- F is **interim implementable** on T if there exists direct mechanism $f : T \to Y$ which is IIC on T, and $f(t) \in F(\hat{\theta}(t))$ for all $t \in T$
- Notion of **robustness** can be built in by requiring F to be interim implementable on large enough type spaces T — i.e., allow beliefs of agents to vary a lot
- Size of the type space determines the restrictiveness of the theory

Different Kinds of Type Spaces

- *Payoff type spaces:* $T_i = \Theta_i$, $\hat{\theta}_i$ is identity map
- Finite type space $\mathcal{T}$ satisfies CPA (with prior p) if there exists $p \in \Delta(\mathcal{T})$ such that $\sum_{t_{-i} \in T_{-i}} p(t_i, t_{-i}) > 0$ for all i, t_i and

$$\hat{\pi}_i(t_i)[t_{-i}] = \frac{p(t_i, t_{-i})}{\sum_{t'_{-i} \in T_{-i}} p(t_i, t'_{-i})}$$

- Standard mechanism design models restrict attention to payoff type spaces with a common prior
- Can require interim implementability on bigger type spaces: e.g., all common prior type spaces, or all type spaces, to capture different notions of robustness

A Simple 'Belief Free' IC Condition

- Consider a specific game where agents are asked to report only their preferences; i should prefer to report payoff type truthfully at any ex post state θ, if all others are reporting truthfully:
- A preference revelation mechanism $f : \Theta \rightarrow Y$ is **ex post IC** if for all i, all $\theta \in \Theta$:

$$u_i(f(\theta), \theta)) \geq u_i(f(\theta_i', \theta_{-i}), \theta)$$

  for all $\theta_i' \in \Theta_i$

- SCC F is **ex post implementable** if there exists $f : \Theta \rightarrow Y$ such that f is ex post IC and $f(\theta) \in F(\theta)$ for all $\theta \in \Theta$

Sufficiency Theorem

Proposition 1 *If F is ex post implementable, it is interim implementable on any type space.*

Proof: Suppose $f^* : \Theta \to Y$ is ex post IC which implements F

Take arbitrary type space $\mathcal{T}$; define direct mechanism on this space $f : \mathcal{T} \to Y$ satisfying $f(t) = f^*(\hat{\theta}(t))$.

Claim f is interim IC on $\mathcal{T}$, i.e.,

$$t_i \in \arg\max_{t_i' \in T_i} \int_{T_{-i}} u_i(f(t_i', t_{-i}), (\hat{\theta}_i(t_i), \hat{\theta}_{-i}(t_{-i}))d\hat{\pi}_i(t_i)[t_{-i}]$$

Proof of Sufficiency, contd.

Maximand equals

$$\int_{T_{-i}} u_i(f^*(\hat{\theta}(t_i'), \hat{\theta}_{-i}(t_{-i})), (\hat{\theta}_i(t_i), \hat{\theta}_{-i}(t_{-i}))d\hat{\pi}_i(t_i)[t_{-i}]$$

Fix any $\theta_{-i} \in \Theta_{-i}$, and let $P_i(t_i)[\theta_{-i}]$ denote probability assessed by t_i that others' preference parameter is θ_{-i}

Then maximand equals

$$\sum_{\theta_{-i} \in \Theta_{-i}} P_i(t_i)[\theta_{-i}] u_i(f^*(\hat{\theta}_i(t_i^{'}), \theta_{-i}); \hat{\theta}_i(t_i), \theta_{-i})$$

Result now follows from ex post IC property of f^*.

Converse: Necessity

Proposition 2 *Suppose the environment is separable. If F is interim implementable on every common prior payoff type space, it is ex post implementable.*

Proof: Consider the type space where it is common knowledge that types of all agents other than i is θ_{-i}, but t_i is observed only by i.

Then interim implementability on this type space implies existence of a function $g^{i,\theta_{-i}}(\theta_i' : \Theta_i \to Y_0 \times Y_1 \times \ldots$ such that IIC condition for i holds, and $g^{i,\theta_{-i}}(\theta_i)(\theta_i) \in F(\theta)$.

Separability of the environment implies
$g_0^{i,\theta_{-i}}(\theta_i)(\theta_i) = f_0(\theta_i, \theta_{-i}), g_j^{i,\theta_{-i}}(\theta_i)(\theta_i) \in F_j(\theta_i, \theta_{-i})$

Finally construct $f(\theta) = (f_0(\theta), \ldots, g_i^{i,\theta_{-i}}(\theta_i), ..)$ and use ex post IC property. $\square$

Corollary to Prop 2

In a separable environment, the following are equivalent conditions:

- F is interim implementable on all type spaces
- F is interim implementable on all common prior type spaces
- F is interim implementable on all payoff type spaces
- F is interim implementable on all common prior payoff type spaces
- F is ex post implementable