



OPTIMAL MECHANISM DESIGN WITH APPROXIMATE INCENTIVE COMPATIBILITY AND MANY PLAYERS

Pathikrit Basu

Freelance Researcher

pathikritbasu@gmail.com

ABSTRACT

We consider a setting in which a mechanism designer must choose the appropriate social alternative depending on the state of nature. We study the problem of optimal design and demonstrate that a mechanism which allocates resources so as to achieve the social optimum and assigns payments equal to the posterior expected utility of the agent at the social optimum, is an ϵ -optimal mechanism for environments with many players.

Keywords: Mechanism design, incentive compatibility, statistical decision theory

JEL Classification Numbers: D60, D61, D62.

1. INTRODUCTION

ONE of the main aspects of the study of mechanism design is aggregating private information in order to reach a socially optimal objective. Since the agents may benefit from particular social alternatives being chosen, they have a willingness to pay, so far as social choice aligns with their preferences. This allows the mechanism designer to extract revenue in the form of payments from the players in the mechanism while ensuring players have the incentive to participate truthfully (Hurwicz, 1960; Gibbard, 1973; Maskin, 1999; Vickrey, 1961; Clarke, 1971; Groves, 1973; Myerson, 1981; Myerson & Satterthwaite, 1983). This paper provides a construction of an optimal mechanism in a setting with many players. The expected payments of the players from the mechanism have a *what you give is what you get*

interpretation. The problem studied in the paper is of the nature of statistical decision problems (Wald, 1950; Blackwell & Girshick, 1979; Berger, 2013; Ferguson, 2014; Pratt et al., 1995; DeGroot, 2005). The result in the paper hinges on the intuition that for settings with many players, an individual player's opinion (type reported) does not affect the aggregate, achieving approximate incentive compatibility. Hence, the socially optimal mechanism allows the mechanism designer to extract all the surplus. This feature in incentive compatibility relates the paper to social learning and herd formation models (Banerjee, 1992; Bikhchandani et al., 1992; Chamley, 2004; Smith & Sørensen, 2000; Fudenberg et al., 2021). Further, the paper is also related to the models in De Condorcet (1785) and Roberts & Postlewaite (1976). Approximate incentive compatibility in mechanism and market design contexts has also been considered in Azevedo & Budish (2019), Balcan et al. (2019), Epasto et al. (2018) and Lee (2016). The result in the paper may also be viewed as a general result in a standard setting in which optimal payments may be characterized as agents paying their expected value which perhaps interestingly, is in contrast with the externality based payment scheme of Pigou (1920) and VCG payments.

2. MODEL

An *environment* is a tuple $\mathcal{E} = \langle N, \Omega, (S_i)_{i \in N}, A, (u_i)_{i \in N}, \pi_0, \mu \rangle$. The set N is a finite set and it is the set of all players in the environment. The set Ω is the set of possible states of nature, assumed to be finite. For each $i \in N$, the set S_i (finite) is the set of possible signals that player i may receive regarding the true state $\omega \in \Omega$. Finally, the set A (finite) is the set of alternatives. Player $i \in N$ has a state-dependent utility function $u_i : A \times \Omega \rightarrow \mathbb{R}$. In terms of the information structure present in the environment, $\pi_0 \in \Delta(\Omega)$ is a common prior and $\mu = \mu(\cdot | \omega)_{\omega \in \Omega} \subseteq \Delta(S)$ is the state-dependent signal distribution for the players, with the set of all possible signal profiles being $S = \prod_{i \in N} S_i$. We assume that $\pi_0(\omega) > 0$ and $\mu(s | \omega) > 0$ for each $s \in S$ and $\omega \in \Omega$. We denote as $\pi_0 \otimes \mu$, the joint distribution on the set $\Omega \times S$, generated by the prior π_0 and signal distribution μ . For $s_i \in S_i$ and $s \in S$, we write $\pi(s_i)$ to be the posterior belief in $\Delta(\Omega)$ conditional on player i 's signal s_i and $\pi(s)$ to be the posterior belief conditional on the signal profile s i.e. the private signals of all the players in the environment.

A *mechanism* is a tuple (σ, q) , in which $\sigma : S \rightarrow A$ is a social choice

function and $q = (q_i)_{i \in N}$ is a collection of payment functions, the payment function for $i \in N$ is a function $q_i : S \rightarrow \mathbb{R}$.

We now state and provide some more definitions.

Definition 1. Let $\varepsilon > 0$. A mechanism (σ, q) is said to be ε -Bayesian incentive compatible if for each $i \in N$ and $s_i, t_i \in S_i$,

$$\mathbb{E}_{\pi_0 \otimes \mu} [u_i(\sigma(s_i, s_{-i}), \omega) - q_i(s_i, s_{-i}) | s_i] \geq \mathbb{E}_{\pi_0 \otimes \mu} [u_i(\sigma(t_i, s_{-i}), \omega) - q_i(t_i, s_{-i}) | s_i] - \varepsilon.$$

Definition 2. A mechanism (σ, q) is said to be Bayesian individually rational if for each $i \in N$ and $s_i \in S_i$,

$$\mathbb{E}_{\pi_0 \otimes \mu} [u_i(\sigma(s_i, s_{-i}), \omega) - q_i(s_i, s_{-i}) | s_i] \geq 0.$$

For any given mechanism (σ, q) , we define the revenue $Q(\sigma, q)$ from the mechanism as the expected sum of payments derived from the mechanism i.e.

$$Q(\sigma, q) = \mathbb{E}_{\pi_0 \otimes \mu} \left[\sum_{i \in N} q_i(s_i, s_{-i}) \right].$$

We now state the definition of an ε -optimal mechanism.

Definition 3. Let $\varepsilon > 0$. A mechanism (σ', q') is said to be ε -optimal if

1. (σ', q') is ε -Bayesian incentive compatible and Bayesian individually rational.
2. For any other mechanism (σ, q) that is ε -Bayesian incentive compatible and Bayesian individually rational,

$$Q(\sigma', q') \geq Q(\sigma, q).$$

We define the following mechanism (σ^*, q^*) , which is the main mechanism proposed by the paper. It implements the social optimum and prescribes payments that are equal to the posterior expected utility of the agent.

1. For each $s \in S$,

$$\sigma^*(s) \in \arg \max_{a \in A} \sum_{\omega \in \Omega} \pi(s)(\omega) \sum_{i \in N} u_i(a, \omega).$$

2. For each $s \in S$, for each $i \in N$,

$$q_i^*(s) = \sum_{\omega \in \Omega} \pi(s)(\omega) u_i(\sigma^*(s), \omega).$$

We prove the main theorem.

Theorem 1. *Suppose we set A and Ω to be the set of alternatives and the set of states of nature. Suppose X (finite) is a signal space. Let $\pi_0 \in \Delta(\Omega)$ be a common prior. Let $\{v(\cdot|\omega)\}_{\omega \in \Omega} \subseteq \Delta(X)$ be a signal distribution such that $v(x|\omega) > 0$ for each $x \in X$, $\omega \in \Omega$ and $v(\cdot|\omega) \neq v(\cdot|\omega')$ for $\omega \neq \omega'$.*

Let \mathcal{U} be a finite set of utility functions $u : A \times \Omega \rightarrow \mathbb{R}$ such that for each $\lambda \in \Delta(\mathcal{U})$ and for each $\omega \in \Omega$, there exists $a \in A$ (unique maximiser) such that

$$\sum_{u \in \mathcal{U}} \lambda(u) u(a, \omega) > \sum_{u \in \mathcal{U}} \lambda(u) u(b, \omega), \quad (1)$$

for each $b \in A \setminus \{a\}$.

Let $\varepsilon > 0$. Then, there exists $n_0 \in \mathbb{N}$ such that for any environment $\mathcal{E} = \langle N, \Omega', (S_i)_{i \in N}, A', (u_i)_{i \in N}, \pi'_0, \mu \rangle$ satisfying

1. $|N| > n_0$;
2. $\Omega' = \Omega$; $\pi'_0 = \pi_0$; $S_i = X$, for each $i \in N$; $\mu(\cdot|\omega) = v^N(\cdot|\omega)$, for each $\omega \in \Omega$ (the probability measure $v^N(\cdot|\omega)$ is the product probability measure in $\Delta(S)$ with index set N , for each $\omega \in \Omega$);
3. $u_i \in \mathcal{U}$, for each $i \in N$, the mechanism (σ^*, q^*) is ε -optimal.

Proof. For each $\lambda \in \Delta(\mathcal{U})$ and any $\omega \in \Omega$, define the set

$$E(\lambda; \omega) = \left\{ e \in [0, 1] : \text{for each } \pi \in \Delta(\Omega), \text{ if } \pi(\omega) > e, \text{ then} \right. \\ \left. \arg \max_{a \in A} \sum_{\omega' \in \Omega} \pi(\omega') \sum_{u \in \mathcal{U}} \lambda(u) u(a, \omega') = \arg \max_{a \in A} \sum_{u \in \mathcal{U}} \lambda(u) u(a, \omega) \right\}.$$

The above defines a correspondence taking input values $\lambda \in \Delta(\mathcal{U})$ and outputs the set $E(\lambda; \omega) \subseteq [0, 1]$ i.e. $E(\cdot; \omega) : \Delta(\mathcal{U}) \rightrightarrows [0, 1]$, a correspondence

$E(\cdot; \omega)$ for each $\omega \in \Omega$. $E(\lambda; \omega)$ is the set of all values e in $[0, 1]$ such that if a belief π assigns probability greater than e on the state ω , then the optimal action for the mixture utility function $\sum_{u \in \mathcal{U}} \lambda(u)u(a, \omega)$ in state ω is the same as in under the belief π . Given that condition 1 in the statement of the theorem is satisfied, it follows that $\arg \max_{a \in A} \sum_{u \in \mathcal{U}} \lambda(u)u(a, \omega)$ is a singleton as there is a unique maximiser. Further, since the expression $\sum_{\omega' \in \Omega} \pi(\omega') \sum_{u \in \mathcal{U}} \lambda(u)u(a, \omega')$ is continuous (linear) in both λ and π , it follows that the correspondence $E(\lambda; \omega)$ is both upper and lower hemicontinuous in λ . Further, by definition, $E(\lambda; \omega)$ is convex and closed since it is always a closed interval of the form $[e', 1]$. Hence, by applying the theorem of the maximum (Charalambos & Aliprantis, 2013), we may prove that $e(\lambda; \omega) := \min_{e \in E(\lambda; \omega)} e$ is continuous for each $\omega \in \Omega$. Now, let $e = \max_{\omega \in \Omega} \max_{\lambda \in \Delta(\mathcal{U})} e(\lambda; \omega)$. Since it may be argued that $e(\lambda; \omega) < 1$ for each $\lambda \in \Delta(\mathcal{U})$ and $\omega \in \Omega$, it follows that $e < 1$.

Let $\varepsilon > 0$. Then, let $\delta \in (0, 1)$ such that

$$5\delta \max_{u \in \mathcal{U}} \max_{a \in A} \|u(a, \cdot)\| < \varepsilon. \tag{2}$$

Let X^∞ be the space of all sequences in X . For each $\omega \in \Omega$, let $v^\infty(\cdot|\omega)$ be the product probability measure in $\Delta(X^\infty)$. For $x^n \in X^n$, let $\pi(x^n)$ be the posterior belief on the state of nature conditional on signals in x^n . Define the following events in X^∞ , one for each $\omega \in \Omega$,

$$X_\omega^\infty := \{x^\infty \in X^\infty : \lim_{n \rightarrow \infty} \pi(x^n)(\omega) = 1\}.$$

Then, it follows that

$$v^\infty(X_\omega^\infty|\omega) = 1. \tag{3}$$

Given $n \in \mathbb{N}$, define the following sets in X^n .

$$P_\omega^n = \{x^n \in X^n : \forall y \in X, \pi(x^n, y)(\omega) > e\}.$$

$$T^n = \{x^n \in X^n : \forall y, z \in X, \|\pi(x^n, y) - \pi(x^n, z)\| < \delta\}.$$

Then, it follows from (3) that there exists $n_0 \in \mathbb{N}$ such that for each $n \geq n_0$ and for each $\omega \in \Omega$,

$$v^n(P_\omega^n \cap T^n|\omega) \geq 1 - \delta, \tag{4}$$

given the n -fold product measure $\mathbf{v}^n(\cdot|\omega)$ in $\Delta(X^n)$.

The chosen n_0 is the one we pick.

Suppose $\mathcal{E} = \langle N, \Omega, (S_i)_{i \in N}, A, (u_i)_{i \in N}, \pi_0, \mu \rangle$ is an environment that satisfies the properties 1), 2) and 3). Then, we show that the mechanism (σ^*, q^*) is ε -optimal.

We first show that (σ^*, q^*) is ε -Bayesian incentive compatible. Let $i \in N$ and $s_i, t_i \in S_i$. Then, we have that

$$\begin{aligned} \mathbb{E}_{\pi_0 \otimes \mu} [(u_i(\sigma^*(t_i, s_{-i}), \omega) - q_i^*(t_i, s_{-i})) - (u_i(\sigma^*(s_i, s_{-i}), \omega) - q_i^*(s_i, s_{-i})) | s_i] \\ \leq \delta \max_{u \in \mathcal{U}} \max_{a \in A} \|u(a, \cdot)\| + 4\delta \max_{u \in \mathcal{U}} \max_{a \in A} \|u(a, \cdot)\|. \end{aligned}$$

The above inequality follows since for the given conditional expectation, from (4), with probability at least $1 - \delta$, two things happen simultaneously : i) the social optimum does not change with the unilateral deviation from s_i to t_i i.e. $\sigma^*(s_i, s_{-i}) = \sigma^*(t_i, s_{-i})$ hence $u_i(\sigma^*(s_i, s_{-i}), \omega) = u_i(\sigma^*(t_i, s_{-i}), \omega)$ and ii) the change in posterior belief is at most of distance δ i.e. $\|\pi(s_i, s_{-i}) - \pi(t_i, s_{-i})\| < \delta$, hence this means that $q_i^*(s_i, s_{-i}) - q_i^*(t_i, s_{-i}) \leq \delta \max_{u \in \mathcal{U}} \max_{a \in A} \|u(a, \cdot)\|$. Further, with probability at most δ , we get a difference of four terms

$$(u_i(\sigma^*(t_i, s_{-i}), \omega) - q_i^*(t_i, s_{-i})) - (u_i(\sigma^*(s_i, s_{-i}), \omega) - q_i^*(s_i, s_{-i}))$$

that takes a value of at most $4 \max_{u \in \mathcal{U}} \max_{a \in A} \|u(a, \cdot)\|$, by the definition of q^* .

Hence, it follows from (2) that

$$\mathbb{E}_{\pi_0 \otimes \mu} [(u_i(\sigma^*(t_i, s_{-i}), \omega) - q_i^*(t_i, s_{-i})) - (u_i(\sigma^*(s_i, s_{-i}), \omega) - q_i^*(s_i, s_{-i})) | s_i] < \varepsilon,$$

which implies that (σ^*, q^*) is ε -Bayesian incentive compatible.

Next, we show that (σ^*, q^*) is Bayesian individually rational. Let $i \in N$ and $s_i \in S_i$. Then,

$$\begin{aligned} \mathbb{E}_{\pi_0 \otimes \mu} [u_i(\sigma^*(s_i, s_{-i}), \omega) | s_i] &= \mathbb{E}_{s_{-i}} \left[\sum_{\omega \in \Omega} \pi(s_i, s_{-i})(\omega) u_i(\sigma^*(s_i, s_{-i}), \omega) | s_i \right] \\ &= \mathbb{E}_{s_{-i}} [q_i^*(s_i, s_{-i}) | s_i] \\ &= \mathbb{E}_{\pi_0 \otimes \mu} [q_i^*(s_i, s_{-i}) | s_i], \end{aligned} \tag{5}$$

implying that (σ^*, q^*) is Bayesian individually rational.

Finally, we prove that (σ^*, q^*) is ε -optimal. Let (σ, q) be any other mechanism that is ε -Bayesian incentive compatible and Bayesian individually rational. We will show that $Q(\sigma^*, q^*) \geq Q(\sigma, q)$ i.e. $\mathbb{E}_{\pi_0 \otimes \mu} [\sum_{i \in N} q_i^*(s_i, s_{-i})] \geq \mathbb{E}_{\pi_0 \otimes \mu} [\sum_{i \in N} q_i(s_i, s_{-i})]$.

Since (σ, q) is Bayesian individually rational, we have that for each $i \in N$ and $s_i \in S_i$,

$$\mathbb{E}_{\pi_0 \otimes \mu} [u_i(\sigma(s_i, s_{-i}), \omega) | s_i] \geq \mathbb{E}_{\pi_0 \otimes \mu} [q_i(s_i, s_{-i}) | s_i].$$

Hence, taking the unconditional expectation and summing over all players, we get that

$$\mathbb{E}_{\pi_0 \otimes \mu} [\sum_{i \in N} u_i(\sigma(s_i, s_{-i}), \omega)] \geq \mathbb{E}_{\pi_0 \otimes \mu} [\sum_{i \in N} q_i(s_i, s_{-i})].$$

Since, the social choice function σ^* implements the social optimum it follows that

$$\mathbb{E}_{\pi_0 \otimes \mu} [\sum_{i \in N} u_i(\sigma^*(s_i, s_{-i}), \omega)] \geq \mathbb{E}_{\pi_0 \otimes \mu} [\sum_{i \in N} u_i(\sigma(s_i, s_{-i}), \omega)].$$

By applying (5), taking unconditional expectations and summing over all players, we obtain

$$\mathbb{E}_{\pi_0 \otimes \mu} [\sum_{i \in N} q_i^*(s_i, s_{-i})] = \mathbb{E}_{\pi_0 \otimes \mu} [\sum_{i \in N} u_i(\sigma^*(s_i, s_{-i}), \omega)].$$

Hence, it follows by the previous conclusions, that $Q(\sigma^*, q^*) \geq Q(\sigma, q)$. Thus, we have proved that the mechanism (σ^*, q^*) is ε -optimal. \square

With the regard to the above theorem, some remarks are in order. Firstly, by applying standard results on convergence rates for Bayesian posteriors (Ibragimov et al., 1981; Le Cam, 1986; Ghosal et al., 2000), we may further derive a threshold on the number of agents $n_0 = N(\varepsilon, \nu)$ of the order of $O(\frac{1}{\varepsilon^2})$. Secondly, the condition on \mathcal{U} implies that each $u \in \mathcal{U}$ admits a unique maximiser, for each state. The condition would also be satisfied if all agents exhibit the same ordinal preference over alternatives. However, agents could

disagree in their state-dependent ordinal ranking over states and yet, condition (1) may be satisfied. The finiteness of \mathcal{U} is applied in the proof of the theorem to ensure that $\epsilon < 1$ indeed exists as the set $\Delta(\mathcal{U})$ is compact. Of course, in this setting, $\lambda(u)$ is the proportion of agents having utility function u . Hence, it is essentially the normalised welfare weight on u in the term for aggregate social welfare.

We next discuss some examples. Consider a situation involving single object assignment in which there exist different sets of agent $\{N_u\}_{u \in \mathcal{U}}$ (pairwise disjoint), where $N = \cup_{u \in \mathcal{U}} N_u$ and each agent in N_u has utility function u . Further, the set of alternatives is defined as $A = \mathcal{U}$, meaning that the object is assigned to exactly one of $\{N_u\}_{u \in \mathcal{U}}$. Whichever N_u is assigned the object, an agent in N_u derives state-dependent utility according to u and would have an expected payment equal to the expected value of obtaining the object for N_u . Hence, if not obtaining the object has no value, this means the agent does not pay anything in the mechanism. Perhaps interestingly, the theorem application above would only need $\sum_{u \in \mathcal{U}} |N_u|$ to go to infinity and hence we may have that one set of agents is large relative to other sets of agents. For another example, one may set aside optimality and instead consider ϵ -Budget Balanced mechanisms (Myerson & Satterthwaite, 1983), where the expected sum of payments would be close to zero. This would be the case, when the distribution of utilities over agents (i.e. profile of utility functions $(u_i)_{i \in N}$) is such that the expected welfare is close to zero, hence the expected sum of payments (for the given payment scheme) would be close to zero. Hence, we would get a mechanism that would satisfy the properties of being ϵ -BIC, BIR and ϵ -Budget Balanced. This would demonstrate a situation that would prove to be in contrast to Myerson & Satterthwaite (1983).

References

- Azevedo, E. M., & Budish, E. (2019). Strategy-proofness in the large. *Review of Economic Studies*, 86(1), 81–116.
- Balcan, M.-F., Sandholm, T., & Vitercik, E. (2019). Estimating approximate incentive compatibility. *arXiv preprint arXiv:1902.09413*.
- Banerjee, A. V. (1992). A simple model of herd behavior. *Quarterly Journal of Economics*, 107(3), 797–817.
- Berger, J. O. (2013). *Statistical Decision Theory and Bayesian Analysis*. Springer.

- Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5), 992–1026.
- Blackwell, D. A., & Girshick, M. A. (1979). *Theory of Games and Statistical Decisions*. Courier Corporation.
- Chamley, C. P. (2004). *Rational Herds: Economic Models of Social Learning*. Cambridge University Press.
- Charalambos, D., & Aliprantis, B. (2013). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer.
- Clarke, E. H. (1971). Multipart pricing of public goods. *Public Choice*, 17–33.
- De Condorcet, N. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*.
- DeGroot, M. H. (2005). *Optimal Statistical Decisions* (Vol. 82). John Wiley & Sons.
- Epasto, A., Mahdian, M., Mirrokni, V., & Zuo, S. (2018). Incentive-aware learning for large markets. In *Proceedings of the 2018 World Wide Web Conference* (pp. 1369–1378).
- Ferguson, T. S. (2014). *Mathematical Statistics: A Decision Theoretic Approach* (Vol. 1). Academic press.
- Fudenberg, D., Lanzani, G., & Strack, P. (2021). Pathwise concentration bounds for misspecified bayesian beliefs. Available at SSRN 3805083.
- Ghosal, S., Ghosh, J. K., & Van Der Vaart, A. W. (2000). Convergence rates of posterior distributions. *Annals of Statistics*, 500–531.
- Gibbard, A. (1973). Manipulation of voting schemes: a general result. *Econometrica*, 587–601.
- Groves, T. (1973). Incentives in teams. *Econometrica*, 617–631.
- Hurwicz, L. (1960). Optimality and informational efficiency in resource allocation processes. *Mathematical Methods in the Social Sciences*.
- Ibragimov, I., Minskii, H., & Zalmanovich, R. (1981). *Statistical Estimation: Asymptotic Theory* (Vol. 16). Springer.
- Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer.
- Lee, S. (2016). Incentive compatibility of large centralized matching markets. *Review of Economic Studies*, 84(1), 444–463.
- Maskin, E. (1999). Nash equilibrium and welfare optimality. *Review of Economic Studies*, 66(1), 23–38.
- Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research*, 6(1), 58–73.
- Myerson, R. B., & Satterthwaite, M. A. (1983). Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 29(2), 265–281.
- Pigou, A. C. (1920). *Economics of Welfare*. MacMillan.

- Pratt, J. W., Raiffa, H., Schlaifer, R., et al. (1995). *Introduction to Statistical Decision Theory*. MIT press.
- Roberts, D. J., & Postlewaite, A. (1976). The incentives for price-taking behavior in large exchange economies. *Econometrica*, 115–127.
- Smith, L., & Sørensen, P. (2000). Pathological outcomes of observational learning. *Econometrica*, 68(2), 371–398.
- Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance*, 16(1), 8–37.
- Wald, A. (1950). *Statistical Decision Functions*. Wiley.