An Introduction to Mechanism Design

E. Maskin
Harvard University

January 2024

Introduction to Mechanism Design

Theory of Mechanism Design –

"engineering" part of economic theory

- much of economic theory devoted to:
  - understanding existing economic institutions
  - explaining/predicting outcomes that institutions generate
  - positive, predictive

- mechanism design – reverses the direction
  - begins by identifying desired outcomes (goals)
  - asks whether institutions (mechanisms) could be designed to achieve goals
  - if so, what forms would institutions take?
  - normative, prescriptive

For example, suppose

- mother wants to divide cake between 2 children, Alice and Bob
- goal: divide so that each child is happy with his/her portion
  - Bob thinks he has got at least half
  - Alice thinks she has got at least half

  call this *fair division*

- If mother knows that the kids see the cake in same way she does, simple solution:
  - she divides equally (in her view)
  - gives each kid a portion

- But what if, say, Bob sees cake differently from mother?
  - *she* thinks she's divided it equally
  - but *he* thinks piece he's received is smaller than Alice's
- difficulty: mother wants to achieve *fair division*
  - but doesn't have enough information to do this on her own
  - in effect, doesn't know which division is fair

- Can she design a mechanism (procedure) for which outcome will be a fair division?

  (even though she doesn't know what is fair herself ?)

- Age-old problem
  - Lot and Abraham dividing grazing land

Age-old solution:

- have *Bob* divide the cake in two
- have *Alice* choose one of the pieces

Why does this work?

- Bob will divide so that pieces are equal in his eyes
  - if one of the pieces were bigger, then Alice would take that one
- So whichever piece Alice takes, Bob will be happy with other
- And Alice will be happy with her own choice because if she thinks pieces unequal, can take bigger one

Example illustrates key features of mechanism design:

- mechanism designer herself *doesn't know* in advance what outcomes are optimal
- so must proceed indirectly through a mechanism
  - have participants *themselves* generate information needed to identify optimal outcome
- complication: participants don't care about *mechanism designer's goals*
  - have their own *objectives*
- so mechanism must be *incentive compatible*
  - must reconcile social and individual goals

Second Example:

Suppose government wants to sell right (license) to transmit on band of radio frequencies

(real-life issue for many governments)

- several telecommunication companies interested in license
- goal of government: to put transmitting license in hands of company that values it most ("efficient" outcome)
- but government doesn't know how much each company values it (so doesn't know best outcome)

Government could ask each company how much it values license

- but if company thinks its chances of getting license go up when it states higher value, has incentive to *exaggerate* value
- so no guarantee of identifying company that values it most

- government could have
  - each company make a bid for license
  - high bidder wins license
  - winner pays bid
- but this mechanism won't work either
  - companies have incentive to *understate*
- suppose license worth $10m to Telemax, then
  - if Telemax bids $10m and wins, gets
    $10m − $10m = 0
- so Telemax will bid *less* than $10m
- but if all bidders are understating, no guarantee that winner will be company that values license most

Solution:

- every company makes bid for license
- winner is high bidder
- winner pays *second-highest* bid
  - so if 3 bidders and bids are
    $10m, $8m, and $5m,
    winner is company that bids $10m
  - but pays only $8m
- Now company has no incentive to understate
  - doesn't pay bid anyway
  - if understates, may lose license
- Has no incentive to overstate
  - If bids $12m, will now win if other company bids $11m
  - But overpays
- So best to bid *exactly* what license worth
- And winner will be company that values license most

Final Example

Consider society with

- 2 consumers of energy – Alice and Bob

- Energy authority – must choose public energy source
  - gas
  - oil
  - nuclear power
  - coal

Two states of world

state 1 consumers weight future lightly (future relatively unimportant)

state 2 consumers weight future heavily (future relatively important)

Alice – cares mainly about convenience

In state 1: favors gas over oil, oil over coal, and coal over nuclear

In state 2: favors nuclear over gas, gas over coal, and coal over oil

- technical advances expected to make gas, coal, and especially nuclear easier to use in future compared with oil

Bob – cares more about safety

In state 1: favors nuclear over oil, oil over coal, and coal over gas

In state 2: favors oil over gas, gas over coal, and coal over nuclear

- disposal of nuclear waste will loom large
- gas will become safer

State 1		State 2	
Alice	Bob	Alice	Bob
gas	nuclear	nuclear	oil
oil	oil	gas	gas
coal	coal	coal	coal
nuclear	gas	oil	nuclear

- energy authority
  - wants source that makes good compromise between consumers' views
  - so, oil is social optimum in state 1
  - gas is social optimum in state 2
- but suppose authority *does not know* state
  - then doesn't know whether oil or gas better

State 1		State 2	
Alice	Bob	Alice	Bob
gas	nuclear	nuclear	oil
oil	oil	gas	gas
coal	coal	coal	coal
nuclear	gas	oil	nuclear
oil optimal		gas optimal	

- authority could ask Alice or Bob about state
  - but Alice has incentive to say "state 2" *regardless* of truth

    always prefers gas to oil

    gas optimal in state 2
  - Bob always has incentive to say "state 1"

    always prefers oil to gas

    oil optimal state 1

  So, simply asking consumers to reveal actual state too naive a mechanism

State 1		State 2	
Alice	Bob	Alice	Bob
gas	nuclear	nuclear	oil
oil	oil	gas	gas
coal	coal	coal	coal
nuclear	gas	oil	nuclear
social optimum: oil		social optimum: gas	

Authority can have consumers participate in the mechanism given by table

		Bob	
		oil	coal
Alice		nuclear	gas

- Alice – can choose top row or bottom row
- Bob – can choose left column or right column
- outcomes given by table entries
- If state 1 holds
  - Alice will prefer top row if Bob plays left column
  - Bob will always prefer left column
  - so (Alice plays top, Bob plays left) is Nash equilibrium
    - neither participant has incentive to change unilaterally to another strategy
  - In fact, it is *unique* Nash equilibrium
    - so good prediction of what Alice and Bob will do

State 1		State 2	
Alice	Bob	Alice	Bob
gas	nuclear	nuclear	oil
oil	oil	gas	gas
coal	coal	coal	coal
nuclear	gas	oil	nuclear
social optimum: oil		social optimum: gas	

		Bob
		oil	coal
Alice	nuclear	gas

So, in state 1:

- expect that
  - Alice will play top strategy
  - Bob will play left strategy
- outcome is oil
- oil is social optimum

State 1		State 2	
Alice	Bob	Alice	Bob
gas	nuclear	nuclear	oil
oil	oil	gas	gas
coal	coal	coal	coal
nuclear	gas	oil	nuclear
social optimum: oil		social optimum: gas	

		Bob	
		oil	coal
Alice		nuclear	gas

Similarly, in state 2:

- expect that
  - Alice will play bottom strategy
  - Bob will play right strategy
- outcome is gas
- gas is social optimum

State 1		State 2	
Alice	Bob	Alice	Bob
gas	nuclear	nuclear	oil
oil	oil	gas	gas
coal	coal	coal	coal
nuclear	gas	oil	nuclear
social optimum: oil		social optimum: gas	

	Bob: oil	Bob: coal
Alice	oil	coal
Alice	nuclear	gas

- Thus, in *either state*, mechanism achieves social optimum, even though
  - mechanism designer doesn't know the state herself
  - Alice and Bob interested in own ends (not social goal)
- We say that mechanism *implements* the designer's goals (oil in state 1, gas in state 2)

- Have shown you mechanisms in the cake, telecommunication, and energy examples
- But analysis may seem a bit *ad hoc*
- Examples prompt questions:
  - is there a *general* way of determining whether or not a given goal is implementable?
  - if it *is* implementable, can we find a mechanism that implements it?
- Answer: yes to both questions

  see Maskin "Nash Equilibrium and Welfare Optimality," 1977

- Have looked at 3 applications of mechanism design theory
- Many other potential applications

1) International treaty on greenhouse gas emissions
2) Policies to prevent financial crises
3) Design of elections