Mechanism Design Theory: How to Implement Social Goals

E. Maskin
Institute for Advanced Study
and
Princeton University

Nobel Lecture
December 8, 2007

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

Outcome

depends on context

- for a government
  - choice of public goods such as
    - infrastructure (e.g., highways)
    - national security/defense
    - environmental protection
    - public education
- for an electorate
  - candidate to fill public office
- for an auctioneer – selling collection of assets
  - allocation of assets across bidders and corresponding payments by bidders
- for a home buyer and a builder contemplating constructing a house
  - specification of house’s characteristics and builder’s remuneration

Which outcome “desirable” or “optimal” also context-dependent:

- for government
  - public good choice that maximizes “net social surplus”

    (social benefit minus cost)
- for electorate
  - candidate that would beat all others in head-to-head competition
- for auctioneer
  - allocation that puts assets into hands of bidders who value them most
  - allocation that maximizes seller’s revenue from sales
- home buyer and builder
  - deal (house specification and remuneration) for which no other deal is preferred by both buyer and seller

Mechanism designer: the one who chooses the institution (procedure, mechanism, game) that determines outcome

• in public good case:	government
• in political case:	framers of political constitution
• in auction case:	auctioneer
• in house case:	buyer and seller *themselves*

- in public good case, if government knows at *outset* which choice of public goods is optimal,
  - then simple mechanism for achieving it:

    government can pass law mandating that choice
- similarly, if auctioneer knows which bidders value assets most,
  - can simply give assets to those bidders

Problem: government or auctioneer *won't* (ordinarily) *have* this information

- surplus-maximizing choice of public goods depends on citizens' *preferences* over all possible alternative public good choices
  - no special reason why government should know these preferences
- likewise, wouldn't expect auctioneer to know bidders' values for assets
- fundamental difficulty for mechanism designers in general:

don't know optimal outcomes (at outset)

- So have to proceed more *indirectly*

  i.e., to design mechanisms that *themselves* generate this information

- Much of my own work and that of many others has addressed questions:

  When is it possible to design such mechanisms?

  What form do mechanisms take?

  And when is it *not* possible to find such mechanisms?

That it is *ever* possible to design such mechanisms may seem surprising

How can mechanism designer attain optimal outcome without even knowing what it is?

So consider simple concrete 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	
Alice	oil	coal
	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 prefer left column if Alice plays top row

  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

Alice	Bob
gas	nuclear
oil	oil
coal	coal
nuclear	gas

social optimum: oil

State 2

Alice	Bob
nuclear	oil
gas	gas
coal	coal
oil	nuclear

social optimum: gas

	Bob	
Alice	oil	coal
	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

Alice	Bob
gas	nuclear
oil	oil
coal	coal
nuclear	gas

social optimum: oil

State 2

Alice	Bob
nuclear	oil
gas	gas
coal	coal
oil	nuclear

social optimum: gas

	Bob	
Alice	oil	coal
	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

Alice	Bob
gas	nuclear
oil	oil
coal	coal
nuclear	gas

social optimum: oil

State 2

Alice	Bob
nuclear	oil
gas	gas
coal	coal
oil	nuclear

social optimum: gas

	Bob	
Alice	oil	coal
	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)
- More generally, in any given setting, determining
  - *whether* or not mechanism designer's goals can be implemented
  - and, if so, *how*

  are major tasks of mechanism design theory

- Intellectual origins of mechanism design:

  Utopian socialists of 19th century

  - repulsed by evils of capitalism
  - believed they could do better

- More direct influence: Planning Controversy of 1930s
  - O. Lange and A. Lerner

    central planning can replicate and even surpass free markets
  - F. von Hayek and L. von Mises

    strenuously denied this possibility

- Controversy important and fascinating but
  - lacked conceptual precision

    crucial terms like "centralization" and "decentralization" not defined
  - lacked technical apparatus, e.g.,

    game theory

    mathematical programming

    to assess each side's claims

Hurwicz (1960), (1972)

- first to give unambiguous definitions of all important concepts
- first to show how technical tools could obtain clear conclusions about issues in debate

Work inspired by Hurwicz has produced consensus that

- von Hayek and von Mises were correct (i.e., market *is* "best" mechanism) in settings where
  - large number of agents (buyers and sellers)

    so that no single agent has much power
  - no significant "externalities"

    other people's consumption or production of a good does not affect *your* consumption or production
- but better mechanisms than market *are* possible if either assumption violated
  - e.g., when goods are *public* (second assumption violated)

    if some people "consume" national security, *everyone* does

Enormous literature derives from Hurwicz

two branches

- particular highly structured settings
  - public goods
  - auctions
  - contracts
- analysis at a *general* level

My own work has fallen in both categories

- today emphasize general results

Hurwicz introduced notion:

social goals being implemented by mechanism

- saw simple example – choosing optimal energy source
- notion of implementation prompts general questions:

when can social goals be implemented?

if implementable, *what* mechanism will do so?

when can social goals *not* be implemented?

- struggled with these questions in mid-1970s
- after (embarrassingly) long time, realized that *monotonicity* of social goals is key to implementation
  - if social goals are not monotonic, then they are not implemetable
  - if social goals *are* monotonic, then (almost) implementable -- need mild additional condition
- monotonicity of social goals:
  - suppose outcome *a* is optimal outcome in state 1
  - if *a* doesn't fall in anyone's ranking (vis à vis any other outcome) in going from state 1 to 2, then *a remains* optimal in state 2
  - but if *a does* fall in someone's ranking

    then *a* need not remain optimal

Consider example from before:

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	

- optimal outcome in state 1 is oil (according to social goals)
- oil doesn’t remain optimal in state 2
- however, oil *falls* in Alice’s ranking (relative to nuclear and coal)
- so social goals *are* monotonic
  - and implementable (as saw earlier)

Modify example a little

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

- note nuclear is attractive option in state 2
  - although ranked third by Alice, ranked first by Bob
  - so nuclear reasonable social goal in state 2
- however, social goals *not* monotonic
  - oil optimal in state 1
  - oil doesn't fall in either person's ranking in going from state 1 to state 2
  - but oil *not* optimal in state 2
- thus, in modified example, social goals cannot be implemented by *any* mechanism

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

To see *why* social goals not implementable,

- suppose, to contrary, there *is* an implementing mechanism
- in that mechanism
  - Alice will play some strategy s_A in state 1
  - Bob will play some strategy s_B in state 2
  - strategies (s_A, s_B) will result in outcome *oil*
- But Alice and Bob will use *same* strategies (s_A, s_B) in state 2
  - only thing Alice prefers to oil is gas
  - but Alice can't have alternative strategy that leads to gas - - would have used it in state 1
  - so won't deviate from s_A in state 2
  - similarly Bob won't deviate from s_B
- so mechanism leads to oil in state 2
  - doesn't implement social goals after all

We have:

Theorem 1: If social goals are implementable, they must be *monotonic*

- in original example, social goals monotonic and implementable
- not always true
  - examples of monotonic social choice rules that are *not* implementable
- still, if additional mild condition imposed, monotonicity *guarantees* implementability

No veto power

- suppose all individuals − except possibly one − agree that outcome *a* is *best possible* outcome (nothing better)
- then *a* must be optimal
  - i.e., remaining individual can't veto it
- quite weak
  - suppose outcome ↔ distribution of economic goods across individuals
  - then each individual wants all goods for himself
  - so no veto power condition *automatically* satisfied

Theorem 2: Suppose "society" has at least 3 individuals

If social goals satisfy monotonicity and no veto power, then implementable

- proof too complicated to present here
  - *constructive*: given social goals, recipe given for explicitly designing mechanism
- Why at least 3 individuals?
  - earlier example had 2 people
  - but implementation, in general, more difficult for 2 than for 3 or more people
  - mechanism
    - gives people incentive to do what they ought to do
    - "punishes" individual for deviating
    - if only 2 people and one has deviated
      - may be hard to tell who has deviated and who hasn't
    - problem resolved with 3 or more people: deviator sticks out

Conclusions

- very brief introduction to mechanism design theory
- of course, much, much more to it
  - other facets in Leo's and Roger's talks
- attraction for me: theory intellectually engaging
  - and also socially useful
- remains lively
  - almost half century after Hurwicz (1960), still active and important part of economic theory
- will be interesting to see where it goes next!