Mechanism Design

9: Information Design

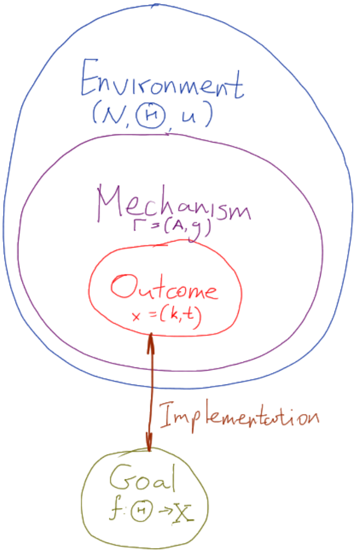
Egor Starkov

Københavns Unversitet
Fall 2023

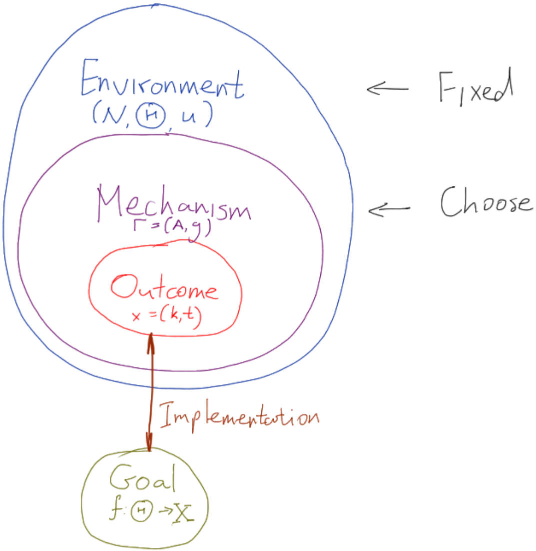
This slide deck:

What is Information Design?

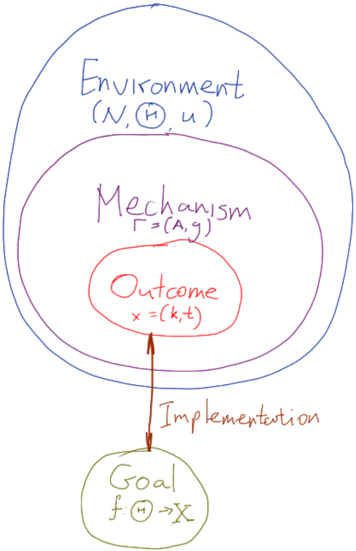
recall our egg diagram...


Environment
(N, Θ, u)
Mechanism
Γ = (A, g)
Outcome
x = (k, t)
Implementation
Goal
f: Θ → X

in Mechanism Design:



in Information Design:



← Choose players' info (Θ)

← Game fixed

What is Information Design?

- In Mechanism design, we have:
  - fixed information and set of outcomes;
  - choose the game that agents play (available actions and their mapping to outcomes).
- Information design (a.k.a. “Bayesian Persuasion”) flips the problem completely:
  - now have a fixed game (actions and outcomes);
  - can choose the information that players have about their payoffs (and others’ payoffs too!)
- Alternatively, in the context of “how to extract information from the player?” problem:
  - We’ve looked at settings when there’s a fixed game between sender(s) and receiver (cheap talk, disclosure) and when receiver can design and commit to various incentive schemes.
  - Now: effectively a communication game, where the *sender* can credibly commit to a certain communication strategy.
- This lecture very vaguely follows Bergemann and Morris [2019]. Kamenica [2019] may also be useful.

Setting

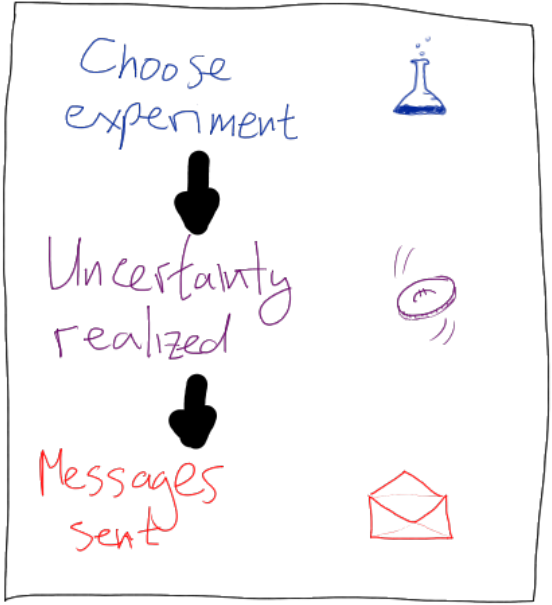
- There is some state $\omega \in \Omega$, unknown to *everyone* initially, common prior $\phi_0 \in \Delta(\Omega)$.
- Players/receivers $i \in \{1, ..., N\}$ (we will mostly look at $N = 1$);
  - each player chooses an action $a_i \in A_i$;
  - player's utility function is $v_i(a, \omega)$, where $a = (a_1, ..., a_N)$.
  - To be clear: both $(A_1, ..., A_N)$ and $(v_1, ..., v_N)$ are **fixed** by the problem environment.
- Designer/sender's objective function is $v_0(a, \omega)$.
- Designer chooses an experiment (μ, M) where $\mu : \Omega \to \Delta(M^N)$.
  (continued on the next slide)

Experiments

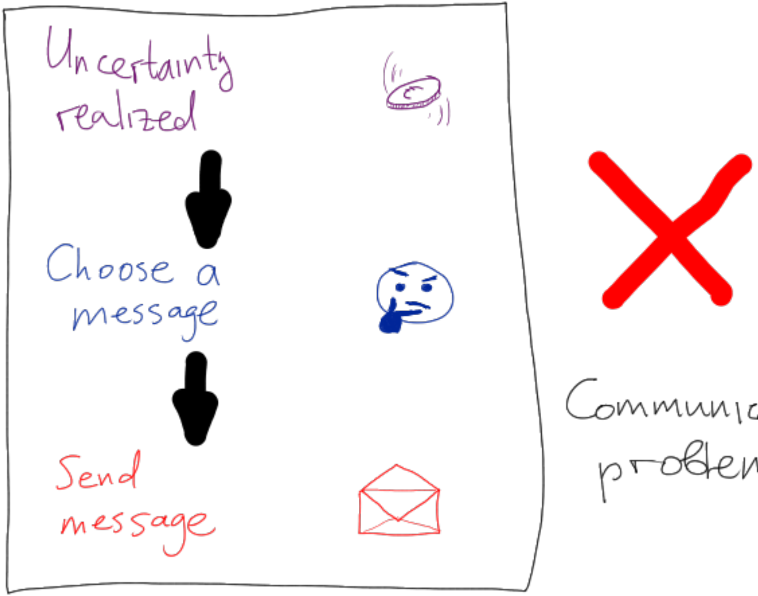
- Designer chooses an experiment $\mu : \Omega \to \Delta(M^N)$.
- In words: an experiment produces, given state ω, some distribution over messages m_i for every player i.
  - For a given ω, $\mu(\omega)$ is a distribution over messages = a (mixed) messaging strategy. So the whole μ prescribes such a messaging strategy for every state ω.
  - Without commitment, $\mu(\omega)$ must be optimal for the sender given ω. That would be a cheap talk model.
- Player i observes message m_i generated by the experiment and uses it to update their belief ϕ about the state, which affects which action a_i they will choose.
- The designer chooses an experiment to manipulate players' info $\to$ players' beliefs $\to$ players' actions.

Experiments

- **Examples**:
  - *Perfectly revealing signals*: (Ω arbitrary and) $m_i = \omega$.
  - *Pooling signals*: ($\Omega \subseteq \mathbb{R}$ and) $m_i = 1$ if $\omega \geq 0$ and $m_i = 0$ if $\omega < 0$.
  - *Partially informative signals*: ($\Omega = \{L, R\}$ and) $m_i = C$ if $\omega = L$ and $m_i = \begin{cases} C & \text{w.p. } 1/2; \\ R & \text{w.p. } 1/2 \end{cases}$ if $\omega = R$.
  - *Uninformative signals*: $m_i \sim F(M)$ where c.d.f. F is *independent* of ω.
- Distinction is made between private persuasion where the designer can send a private signal to each player and public persuasion where only public signals are available.
  - Public signals are available under private persuasion, but not vice versa.
  - Private persuasion is thus always weakly better for the designer.
- Further, timing is important...



Persuasion problem



Communication problem

Experiments

- It is crucial that the designer does not know ω when choosing μ.
- Like in mechanism design, the designer publicly announces and commits to μ.
  I.e., can promise ex ante to reveal something not worth revealing ex post:
  - e.g., commit to telling all truth and nothing but truth – even if unfavorable facts may come up;
  - or commit to giving no hints to players – even if really want to reveal the state sometimes.
- Without commitment – if designer chooses a message *after* learning ω – we have a communication/cheap talk problem (rather than information design problem).
  - These are much more difficult to analyze, mostly due to linkages across states.
  - E.g. a manager wants to boast high earnings to investors but does not want to reveal if earnings low. Without commitment, always reveals when earnings are high. Then if investors hear nothing, they infer that earnings are low – communication in high-earning state imposes an informational externality on low-earning state.

What are “experiments”?

Such a commitment to an experiment/communication strategy can be maintained via:

- **hardware/software**: the sender hardcodes the comm strategy into a device that observes the state and sends a corresponding message (example next slide)
- **blockchain smart contract**: special case of the software commitment above
- **reputation**: if sender & receiver interact repeatedly, receiver can punish sender for deviating from an announced strategy if such deviations observable [Best and Quigley, 2020]

Where are "experiments"? (1)

IRL **example** of such "experiments" (commitment to comm strategies) [Bondi et al., 2020]:

- Wildlife poaching is a problem in many natural preserves. South Africa uses aerial drones with IR/video cameras to (1) detect poachers, (2) alert rangers in the vicinity, if any, and (3) alert poachers they've been detected, for deterrence.
- In this setting: $\omega \in \{\text{detected, undetected}\} \times \{\text{rangers coming, rangers not coming}\}$
- Messages (to poachers) are $m \in \{\text{alert, no alert}\}$. Messaging strategy encoded in software = commitment.
- Conflict: want to let poachers know they are detected and rangers are coming in order to make them flee, but want the threat to be credible (i.e., not signal too often when no rangers are actually in range).
- The idea (tradeoff between deterrence and credibility) is common to "security games": highway patrols, accounting audits, hacking detection, ...

Where are “experiments”? (2)

Other **examples** of such problems from Kamenica [2019] survey (refer to it for exact references):

- financial sector stress tests (Goldstein & Leitner 2018, Inostroza & Pavan 2018, Orlov et al. 2018b),
- grading in schools (Boleslavsky & Cotton 2015, Ostrovsky & Schwarz 2010),
- employee feedback (Habibi 2018, Smolin 2017),
- law enforcement deployment (Hernandez & Neeman 2017, Lazear 2006, Rabinovich et al. 2015),
- censorship (Gehlbach & Sonin 2014),
- entertainment (Ely et al. 2015),
- financial over-the-counter markets (Duffie et al. 2017),
- voter coalition formation (Alonso & Camara 2016b),
- research procurement (Yoder 2018),
- contests (Feng & Lu 2016, Zhang & Zhou 2016),
- medical testing (Schweizer & Szech 2019),
- medical research (Kolotilin 2015),
- matching platforms (Romanyuk & Smolin 2019),
- price discrimination (Bergemann et al. 2015),
- financing (Szydlowski 2016),
- insurance (Garcia & Tsur 2018),
- transparency in organizations (Jehiel 2015),
- routing software (Das et al. 2017, Kremer et al. 2014).

Two interpretations of Information Design

- There are two schools of thought in InfoDesign literature.
- The first takes the literal interpretation, as presented above:
  - there is some designer, who decides how to provide information.
  - School led by Kamenica, most applications take this stand.
- The alternative is a metaphorical interpretation:
  - Suppose we are looking at some game or real-world interaction, but do not know what information is available to players.
  - InfoDesign tools allow to describe the full set of possible outcomes for all possible information structures.
  - No explicit designer in this story.
  - Note that in this story all players have common knowledge of the information structure in place, it is only us (the external observer) who do not know it. (Although there are sophisticated epistemologic arguments for why this is without loss.)
  - Agenda pushed heavily by Morris.

This slide deck:

Illustrative example

The Witness

- A suspect is on trial, accused of murder.
- Judge must decide whether to convict or acquit him, wants to make the right decision.
- Prosecutor is paid per cases won, so wants to convict the suspect regardless of guilt.
- Prosecutor can call up a witness. What kind of witness should he summon?

Illustrative example

The Witness

- A suspect is on trial, accused of murder.
- Judge must decide whether to convict or acquit him, wants to make the right decision.
- Prosecutor is paid per cases won, so wants to convict the suspect regardless of guilt.
- Prosecutor can call up a witness. What kind of witness should he summon?

Frame the above as an information design problem.

Decyphering the example

- Designer $=$ prosecutor.
- State $\omega \in \{G, N\}$ represents true guilt.
- Let $\phi_0 = \mathbb{P}\{\omega = G\}$ denote the common prior belief (probability that the prosecutor and the judge assign to the suspect being guilty).
- $N = 1$ (judge), $A = \{g, n\}$ (verdicts "guilty", "not guilty");

Decyphering the example

- Designer = prosecutor.
- State $\omega \in \{G, N\}$ represents true guilt.
- Let $\phi_0 = \mathbb{P}\{\omega = G\}$ denote the common prior belief (probability that the prosecutor and the judge assign to the suspect being guilty).
- $N = 1$ (judge), $A = \{g, n\}$ (verdicts "guilty", "not guilty");
- The judge's utility is $v_1(a, \omega) = \mathbb{I}(a = \omega)$.
  - You can model it in an asymmetric way too (convicting the innocent can be more or less costly than letting a criminal go).
- The prosecutor's objective function is $v_0(a, \omega) = \mathbb{I}(a = G)$.
  - Prefers the "guilty" verdict regardless of state.

Decyphering the example

- Designer = prosecutor.
- State $\omega \in \{G, N\}$ represents true guilt.
- Let $\phi_0 = \mathbb{P}\{\omega = G\}$ denote the common prior belief (probability that the prosecutor and the judge assign to the suspect being guilty).
- $N = 1$ (judge), $A = \{g, n\}$ (verdicts "guilty", "not guilty");
- The judge's utility is $v_1(a, \omega) = \mathbb{I}(a = \omega)$.
  - You can model it in an asymmetric way too (convicting the innocent can be more or less costly than letting a criminal go).
- The prosecutor's objective function is $v_0(a, \omega) = \mathbb{I}(a = G)$.
  - Prefers the "guilty" verdict regardless of state.
- The witness was at a certain place on the night of murder – this determines μ
  - If W was around the place of murder, can confirm or deny the suspect was there.
  - if W was in a random pub, can do the same, but this conveys different information.

Example: timing

- To be clear, the timing in this example (as well as in the general model) is as follows:
  1. state ω is determined, observed by no one
  2. prosecutor chooses the witness μ and publicly commits to it
  3. witness reveals message m to the court according to $\mu(m|\omega)$
  4. the judge observes m and chooses decision a
  5. payoffs are realized

Example: 4: actions

- Start from the end (proceed by backwards induction):
- Let ϕ denote the judge's posterior belief (after she observes m). What action does she choose?
- Denote $\hat{a}(\phi) \equiv \arg\max \mathbb{E}_{\phi(\omega)}[v_1(a, \omega)]$. If there are many optimal actions, choose the best for the prosecutor.
  - For the first time ever we want to fix the tie-breaking rule. The reason will be evident later.

Example: 4: actions

- Start from the end (proceed by backwards induction):
- Let ϕ denote the judge's posterior belief (after she observes m). What action does she choose?
- Denote $\hat{a}(\phi) \equiv \arg\max \mathbb{E}_{\phi(\omega)}[v_1(a, \omega)]$. If there are many optimal actions, choose the best for the prosecutor.
  - For the first time ever we want to fix the tie-breaking rule. The reason will be evident later.
- In our example:

$$\hat{a}(\phi) = \begin{cases} g & \text{if } \phi \geq 1/2; \\ n & \text{if } \phi < 1/2. \end{cases}$$

Example: 3d: posteriors

- Knowing $\hat{a}(\phi)$ means we can write the prosecutor's utility as a function of ϕ: let

$$V_0(\phi) \equiv v_0(\hat{a}(\phi)) = \begin{cases} 1 & \text{if } \phi \geq 1/2; \\ 0 & \text{if } \phi < 1/2. \end{cases}$$

Example: 3d: posteriors

- Knowing $\hat{a}(\phi)$ means we can write the prosecutor's utility as a function of ϕ: let

$$V_0(\phi) \equiv v_0(\hat{a}(\phi)) = \begin{cases} 1 & \text{if } \phi \geq 1/2; \\ 0 & \text{if } \phi < 1/2. \end{cases}$$

- By choosing an experiment (μ, M) the prosecutor induces some distribution τ over posteriors ϕ. **Trick**: forget about μ and focus on this distribution τ as the choice object
  - What if P could choose any distribution? Would want $\phi \geq 1/2$ always (after any message m).
  - So if $\phi_0 \geq 1/2$ then optimal for P to do nothing (choose uninformative experiment).
  - But the ideal is unattainable if $\phi_0 < 1/2$ because beliefs must be consistent.

Example: 3d: posteriors

- Knowing $\hat{a}(\phi)$ means we can write the prosecutor's utility as a function of ϕ: let

$$V_0(\phi) \equiv v_0(\hat{a}(\phi)) = \begin{cases} 1 & \text{if } \phi \geq 1/2; \\ 0 & \text{if } \phi < 1/2. \end{cases}$$

- By choosing an experiment (μ, M) the prosecutor induces some distribution τ over posteriors ϕ. **Trick**: forget about μ and focus on this distribution τ as the choice object
  - What if P could choose any distribution? Would want $\phi \geq 1/2$ always (after any message m).
  - So if $\phi_0 \geq 1/2$ then optimal for P to do nothing (choose uninformative experiment).
  - But the ideal is unattainable if $\phi_0 < 1/2$ because beliefs must be consistent.
- Belief consistency: $\mathbb{E}_\mu \phi = \phi_0$ (Law of iterated expectations).
  Expectation is taken from the ex ante perspective.
  - Remark: always keep track of perspective. In ID you have ex ante expectations, expectations conditional on m, expectations conditional on ω – easy to get lost!

Example: 3c: feasible distributions

- Let us try to find the best (for P) distribution of posteriors ϕ such that $\mathbb{E}\phi = \phi_0 < 1/2$.
  - It only matters for $V_0(\phi)$ whether $\phi < 1/2$ or $\phi \geq 1/2$.
  - So suppose there are two possible posteriors induced by the experiment: $\phi_1 < 1/2$ and $\phi_2 \geq 1/2$, occurring with respective probabilities τ_1 and $\tau_2 = 1 - \tau_1$.
  - Consistency pins the probabilities exactly:

$$\tau_1\phi_1 + (1-\tau_1)\phi_2 = \phi_0$$
$$\Leftrightarrow \tau_1 = \frac{\phi_2 - \phi_0}{\phi_2 - \phi_1} = 1 - \frac{\phi_0 - \phi_1}{\phi_2 - \phi_1} \tag{1}$$

    (Note that this also implies that $\phi_1 < \phi_0$, since must have $\tau_1 \in [0, 1]$.)
  - (This is not the optimal distribution yet, we only computed τ *given* ϕ_1, ϕ_2 – but we also want to find the optimal values of ϕ_1, ϕ_2)

Example: 3b: optimal distribution

- P gets payoff 1 whenever ϕ_2 is induced and 0 in case of ϕ_1:

$$\mathbb{E}_\phi V_0(\phi) = \tau_1 \cdot 0 + (1 - \tau_1) \cdot 1 = 1 - \tau_1$$

  - Hence want to choose (ϕ_1, ϕ_2) so as to minimize τ_1 subject to (1), $\phi_1 \in [0, \phi_0)$, and $\phi_2 \in [1/2, 1]$.
  - The solution is $\phi_1 = 0$, $\phi_2 = 1/2$. (The objective function is increasing in both ϕ_1 and ϕ_2, so FOCs never hold – thus you only need to check the edges of the domain.)
- So the optimal distribution is: induce posterior $\phi_1 = 0$ with probability $\tau_1 = 1 - 2\phi_0$ and posterior $\phi_2 = 1/2$ with probability $\tau_2 = 2\phi_0$.
  - This yields utility $2\phi_0$ to the designer.

Example: 3a: from distribution to experiment

- Final part: how to design an experiment (μ, M) to induce the desired distribution τ?
  - (It can be shown that this problem always has a solution if τ is *consistent* with the prior ϕ_0.)
  - Have two posteriors so use two messages: $M = \{g, n\}$, where n will induce posterior ϕ_1, and g corresponds to ϕ_2.
  - Let $p(m|\omega)$ denote the probability that message m is sent in state ω. We then need to solve the following system w.r.t. $p(m|\omega)$:

$$\frac{\phi_0 p(n|G)}{\phi_0 p(n|G) + (1-\phi_0)p(n|N)} = \phi_1 = 0$$
$$\frac{\phi_0 p(g|G)}{\phi_0 p(g|G) + (1-\phi_0)p(g|N)} = \phi_2 = 1/2$$
$$p(n|N) + p(g|N) = 1$$
$$p(n|G) + p(g|G) = 1$$

First two equalities ensure that messages b and g induce exactly the right posteriors (and follow from Bayes' rule); the last two are feasibility constraints (one of the two messages must always be sent).

Example: optimal experiment

- **The solution** is given by:

$$p(g|N) = \frac{\phi_0}{1-\phi_0} = 1 - p(n|N)$$
$$p(g|G) = 1 = 1 - p(n|G)$$

  - i.e., in state $\omega = G$ always send $m = g$;
  - in state $\omega = N$ mix between sending $m = g$ w.p. $\frac{\phi_0}{1-\phi_0}$ and $m = n$ w.p. $\frac{1-2\phi_0}{1-\phi_0}$,
- In other words, the optimal strategy is:
  - if state favorable to prosecutor then disclose it truthfully;
  - if state bad for prosecutor then try to obfuscate it.
  - Need commitment to mix in $\omega = N$: message $m = g$ gives higher payoff, so without commitment the prosecutor would never send $m = n$.
- The judge is granted full confidence when taking action that is undesirable for designer; is made barely indifferent when taking action desired by the designer.

This slide deck:

Setting (reminder)

- There is some state $\omega \in \Omega$, unknown to everyone initially, common prior $\phi_0 \in \Delta(\Omega)$.
- Players $i \in \{1, ..., N\}$ (we will mostly look at $N = 1$);
  - each player chooses an action $a_i \in A_i$;
  - player's utility function is $v_i(a, \omega)$, where $a = (a_1, ..., a_N)$.
- Designer's objective function is $v_0(a, \omega)$.
- Designer chooses an experiment (μ, M) where $\mu : \Omega \to \Delta(M^N)$.

Timing

1. Experiment (μ, M) is selected by the designer.
2. Message m is generated by the experiment according to $\mu(m|\omega)$ (where ω is the true realized state). Every player i updates beliefs.
   - Given message m_i, the probability that i's posterior belief assigns to any ω is given by:

$$\phi_i(\omega|m_i) = \frac{\mu(m_i|\omega)\phi_0(\omega)}{\sum_{\omega' \in \Omega} \mu(m_i|\omega')\phi_0(\omega')}$$

Timing

1. Experiment (μ, M) is selected by the designer.
2. Message m is generated by the experiment according to $\mu(m|\omega)$ (where ω is the true realized state). Every player i updates beliefs.
   - Given message m_i, the probability that i's posterior belief assigns to any ω is given by:

$$\phi_i(\omega|m_i) = \frac{\mu(m_i|\omega)\phi_0(\omega)}{\sum_{\omega' \in \Omega} \mu(m_i|\omega')\phi_0(\omega')}$$

   - Note that Belief consistency is then a direct consequence of the law of total probability. Pick some ω and calculate the expected (over m_i) probability that ϕ_i will assign to it:

$$\mathbb{E}_{\mu}[\phi_i(\omega|m_i)|\phi_0] = \sum_{m_i \in M_i} \phi_i(\omega|m_i) \cdot \left[\sum_{\omega' \in \Omega} \mu(m_i|\omega')\phi_0(\omega') \right]$$

Timing

1. Experiment (μ, M) is selected by the designer.
2. Message m is generated by the experiment according to $\mu(m|\omega)$ (where ω is the true realized state). Every player i updates beliefs.
   - Given message m_i, the probability that i's posterior belief assigns to any ω is given by:

$$\phi_i(\omega|m_i) = \frac{\mu(m_i|\omega)\phi_0(\omega)}{\sum_{\omega' \in \Omega} \mu(m_i|\omega')\phi_0(\omega')}$$

   - Note that Belief consistency is then a direct consequence of the law of total probability. Pick some ω and calculate the expected (over m_i) probability that ϕ_i will assign to it:

$$\begin{aligned}\mathbb{E}_\mu[\phi_i(\omega|m_i)|\phi_0] &= \sum_{m_i \in M_i} \phi_i(\omega|m_i) \cdot \left[\sum_{\omega' \in \Omega} \mu(m_i|\omega')\phi_0(\omega')\right] \\ &= \sum_{m_i \in M_i} \frac{\mu(m_i|\omega)\phi_0(\omega)}{\sum_{\omega' \in \Omega} \mu(m_i|\omega')\phi_0(\omega')} \cdot \left[\sum_{\omega' \in \Omega} \mu(m_i|\omega')\phi_0(\omega')\right]\end{aligned}$$

Timing

1. Experiment (μ, M) is selected by the designer.
2. Message m is generated by the experiment according to $\mu(m|\omega)$ (where ω is the true realized state). Every player i updates beliefs.
   - Given message m_i, the probability that i's posterior belief assigns to any ω is given by:

$$\phi_i(\omega|m_i) = \frac{\mu(m_i|\omega)\phi_0(\omega)}{\sum_{\omega' \in \Omega} \mu(m_i|\omega')\phi_0(\omega')}$$

   - Note that Belief consistency is then a direct consequence of the law of total probability. Pick some ω and calculate the expected (over m_i) probability that ϕ_i will assign to it:

$$\begin{aligned}\mathbb{E}_\mu[\phi_i(\omega|m_i)|\phi_0] &= \sum_{m_i \in M_i} \phi_i(\omega|m_i) \cdot \left[\sum_{\omega' \in \Omega} \mu(m_i|\omega')\phi_0(\omega')\right] \\ &= \sum_{m_i \in M_i} \frac{\mu(m_i|\omega)\phi_0(\omega)}{\sum_{\omega' \in \Omega} \mu(m_i|\omega')\phi_0(\omega')} \cdot \left[\sum_{\omega' \in \Omega} \mu(m_i|\omega')\phi_0(\omega')\right]\end{aligned}$$

Timing

1. Experiment (μ, M) is selected by the designer.
2. Message m is generated by the experiment according to $\mu(m|\omega)$ (where ω is the true realized state). Every player i updates beliefs.
   - Given message m_i, the probability that i's posterior belief assigns to any ω is given by:

$$\phi_i(\omega|m_i) = \frac{\mu(m_i|\omega)\phi_0(\omega)}{\sum_{\omega'\in\Omega}\mu(m_i|\omega')\phi_0(\omega')}$$

   - Note that Belief consistency is then a direct consequence of the law of total probability. Pick some ω and calculate the expected (over m_i) probability that ϕ_i will assign to it:

$$\begin{aligned}\mathbb{E}_\mu[\phi_i(\omega|m_i)|\phi_0] &= \sum_{m_i\in M_i}\phi_i(\omega|m_i)\cdot\left[\sum_{\omega'\in\Omega}\mu(m_i|\omega')\phi_0(\omega')\right]\\ &= \sum_{m_i\in M_i}\frac{\mu(m_i|\omega)\phi_0(\omega)}{\sum_{\omega'\in\Omega}\mu(m_i|\omega')\phi_0(\omega')}\cdot\left[\sum_{\omega'\in\Omega}\mu(m_i|\omega')\phi_0(\omega')\right]\\ &= \sum_{m_i\in M_i}\mu(m_i|\omega)\phi_0(\omega) = \phi_0(\omega)\end{aligned}$$

Timing

1. Experiment (μ, M) is selected by the designer.
2. Message m is generated by the experiment according to $\mu(m|\omega)$ (where ω is the true realized state). Every player i updates beliefs.
   - Given message m_i, the probability that i's posterior belief assigns to any ω is given by:

     $$\phi_i(\omega|m_i) = \frac{\mu(m_i|\omega)\phi_0(\omega)}{\sum_{\omega' \in \Omega} \mu(m_i|\omega')\phi_0(\omega')}$$

   - In general, every i also needs to form beliefs over others' messages(="types"), since m_j determine ϕ_j and a_j.
     - Step only relevant if: (1) $N > 1$ and (2) private persuasion is available to designer.
     - We will not look carefully at this step.
3. Every i selects optimal/equilibrium action $\hat{a}_i$ given their beliefs.

Actions are irrelevant

- Restrict attention to public experiments
  - so all players always receive same message m and end up with same posterior belief $\phi \in \Delta(\Omega)$.
  - I do not think this general approach can easily account for private experiments.
- Let $\hat{a}(\phi)$ denote an equilibrium (BNE) strategy profile in a game where all i share belief ϕ.
  - Same as in the example, except now want equilibrium rather than just maximizer.
  - If many equilibria, select designer-best.
- Define $V_0(\phi) \equiv \mathbb{E}_\phi\left[v_0(\hat{a}(\phi), \omega)\right]$.
  - Again, posterior ϕ completely defines the designer's objective function.
  - Expectation needed because v_0 depends on ω in general.
  - Expectation is w.r.t posterior ϕ. Why not prior ϕ_0?

Distributions over posteriors

- As we saw, any experiment (μ, M) induces a distribution over posteriors ϕ. Denote it as $\tau \in \Delta(\Delta(\Omega))$.
  - Under public persuasion, τ is the same for all players.
  - In particular, the [unconditional] probability of posterior ϕ occurring under μ is:

$$\tau(\phi) = \begin{cases} \sum_{\omega' \in \Omega} \mu(m|\omega')\phi_0(\omega') & \text{where } m \text{ is s.t. } \phi(\omega|m) = \phi; \\ 0 & \text{if no such } m \text{ exists.} \end{cases}$$

- Remember belief consistency: $\mathbb{E}_\mu \phi = \sum_{m \in M} \phi \tau(\phi) = \phi_0$.

Direct Experiments

- Note that messages' only purpose is to induce some posterior. So we can w.l.o.g. focus on experiments which directly tell the player what posterior she must have upon hearing this message:

Definition (Direct Experiment A)

A direct experiment is (μ, M) such that $M = \Delta(\Omega)$.

- For the player to actually arrive at beliefs prescribed by a direct experiment, the experiment must be consistent:

Definition

A direct experiment $(\mu, \Delta(\Omega))$ is consistent if $\mathbb{E}_{\tau} m_k = \phi_0$.

Revelation Principle

Theorem (Revelation Principle A)

- *For any experiment (μ, M) there exists an equivalent credible consistent experiment $(\mu, \Delta(\Omega))$.*
- *For any consistent distribution $\tau \in \Delta(\Delta(\Omega))$ there exists an experiment (μ, M) that induces it.*

- So instead of maximizing V_0 over all possible experiments (μ, M) we can maximize over the set of consistent distributions τ, which is a slightly easier problem.
- This is what we did in the example, but now we know this is a way we can always go.
- From this point onwards, I will also call any consistent distribution τ an experiment.

Concavification

- So how do we find the optimal experiment $\tau^{opt} \in \arg\max_\tau \{\mathbb{E}_\tau V_0(\phi)\}$?

Concavification

- So how do we find the optimal experiment $\tau^{opt} \in \arg\max_{\tau} \{\mathbb{E}_{\tau} V_0(\phi)\}$?
- Using the following scary object:

Definition (Concave closure)

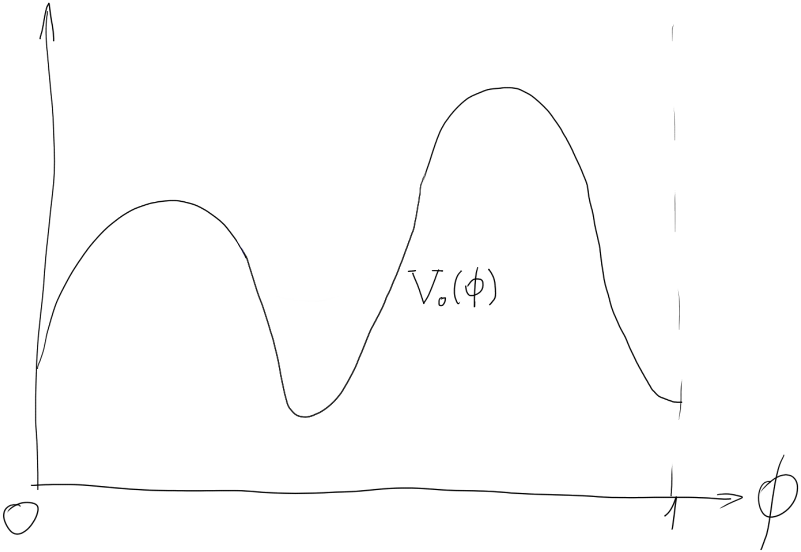
Function $V_0^*(\phi)$ is a concave closure of function $V_0(\phi)$ if it is the pointwise smallest concave function among those that satisfy $V_0^*(\phi) \geq V_0(\phi)$ for all ϕ.

- Equivalent definition:

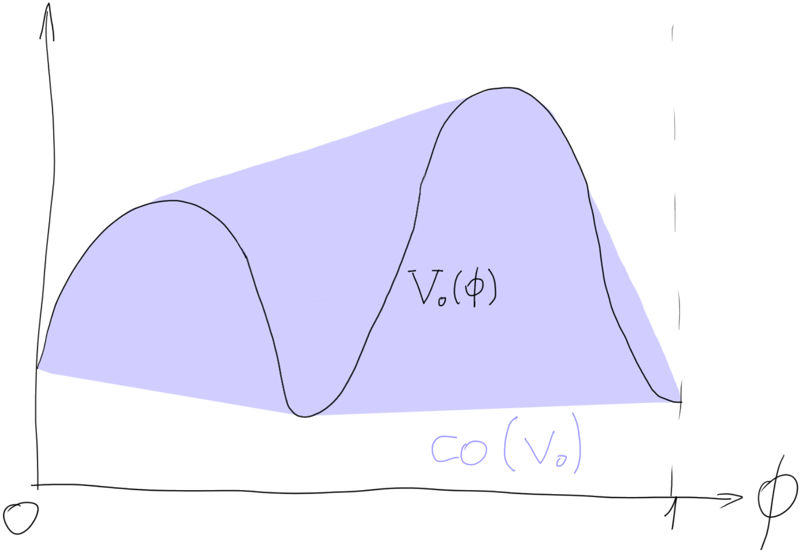
$$V_0^*(\phi) \equiv \sup\{z | (\phi, z) \in co(V_0)\}$$

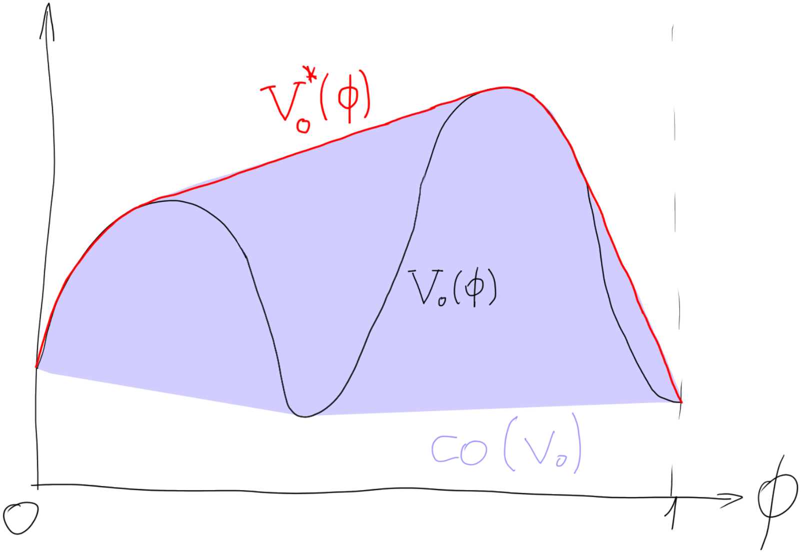
  where $co(V_0)$ is the convex hull of the graph of V_0. V_0^* is then an upper envelope of this convex hull.

- Equivalent definition: V_0^* is the shape that the blanket takes when you throw it over V_0.



$\bar{V}_0(\phi)$
0
1
ϕ


$\bar{V}_0(\phi)$
$co(V_0)$
0
1
ϕ



$V_0^*(\phi)$
$V_0(\phi)$
$co(V_0)$
0
1
ϕ

Optimal experiment: payoff

Theorem (Kamenica & Gentzkow)

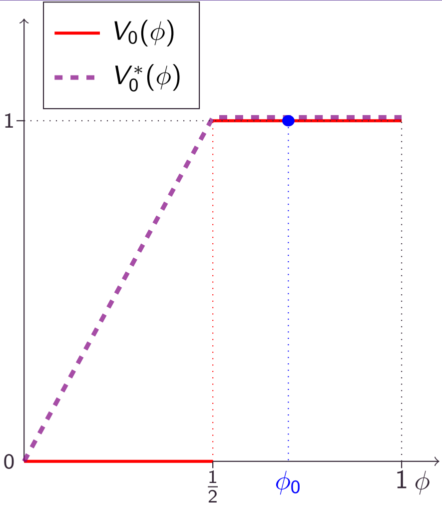
The payoff the designer obtains from the optimal experiment is given by $V_0^*(\phi_0)$, where V_0^* is the concave closure of V_0.

- If $V_0^*(\phi_0) = V_0(\phi_0)$ then trivial (uninformative) experiment is optimal given prior ϕ_0.

Optimal experiment: cookbook

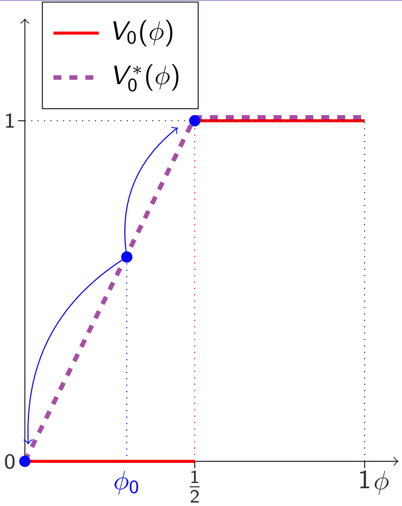
- If $V_0^*(\phi_0) > V_0(\phi_0)$ then you need to find a set of points $\{\phi_1, ..., \phi_K\}$ such that:
  - $\phi_0 \in co(\{\phi_1, ..., \phi_k\})$, meaning $\phi_0 = \sum_k \tau_k \phi_k$ for some weights τ_k ($\tau_k \geq 0$, $\sum_k \tau_k = 1$);
  - $V_0^*(\phi_k) = V_0(\phi_k)$;
  - $V_0^*(\phi_0) = \sum_k \tau_k V_0^*(\phi_k)$.
- These ϕ_k will be the posteriors(=messages) in the optimal experiment τ.
- You can then use all (τ_k, ϕ_k) to derive the conditional probabilities $\mu(m_k|\omega)$ of sending each message m_k in each state ω;
  - use Bayes' rule + feasibility, as we did in the example.
- Congratulations, you have just information designed.

Back to example



- If $\phi_0 \geq 1/2$ then $V_0^*(\phi_0) = V_0(\phi_0)$,
- so trivial mechanism is optimal.

Back to example



- If $\phi_0 < 1/2$ then $V_0^*(\phi_0) > V_0(\phi_0)$.
- Decompose ϕ_0 into points such that $V_0^*(\phi) = V_0(\phi)$, namely $\phi_1 = 0$ and $\phi_2 = 1/2$
- which is exactly what we did when solving the example.

This slide deck:

Introduction

- Kamenica's "concavification" approach yields a nice visual representation of what "Bayesian persuasion" entails.
- But it is not very convenient to use in applications.
  - Concavification is a very graphical concept.
  - Easy to draw a picture for one-dimensional belief ϕ (two states),
  - in principle you can draw one in three dimensions (three states),
  - but with more states things become problematic.
- But there is another, more boring but also more effective approach that links Information Design to an old literature on Correlated Equilibria.
  - You can read more about it in Bergemann & Morris (2019) survey.

Revelation Principle B

- In the previous approach, message m prescribed the posterior belief that a player must update to upon receiving that message.
- Now let it prescribe an action that a player must take:

  A decision rule is $\sigma : \Omega \to \Delta(A)$.
  - Given state ω, give an action recommendation to every player.
  - Recommendations may be random given state, and be correlated across players.
- We will use decision rules as yet another representation of an experiment, along with (μ, M) – message distribution, and τ – distribution of posteriors.
  - Every experiment induces some decision rule. Need to understand which decision rules can be implemented using some experiment.
- Remark: we are no longer constrained to public experiments, now assume that messages are private.

Obedience

Definition

Decision rule σ satisfies obedience if for all i and $a_i, a_i' \in A_i$:

$$\sum_{a_{-i},\omega} v_i((a_i, a_{-i}), \omega)\sigma((a_i, a_{-i})|\omega)\phi_0(\omega) \geq$$

$$\geq \sum_{a_{-i},\omega} v_i((a_i', a_{-i}), \omega)\sigma((a_i, a_{-i})|\omega)\phi_0(\omega)$$

- In words, when i receives a recommendation to play a_i, following it must be better than playing any other a_i' – like our usual IC conditions

Optimal experiment

Theorem (Bergemann & Morris)

A decision rule σ can be induced by an experiment if and only if σ satisfies obedience.

- The designer's problem then is choosing an obedient σ that maximizes

$$v_0^*(\sigma) \equiv \sum_{a,\omega} v_0(a,\omega)\sigma(a|\omega)\phi(\omega)$$

- This is a linear programming problem (both objective and obedience constraints are linear in σ), so trivial to solve in general.
  - At least when Ω and A are finite sets.
- Linear programming is easy, unlike concavification, so can extend this approach easily:
  - already allowed us to look at private messages and not only public;
  - can allow for players' private information (IC constraints also linear);

Back to example

- In the example, the designer's problem is:

$$\begin{aligned} \max_{\sigma} \ & \{\phi_0\sigma(g|G) + (1-\phi_0)\sigma(g|N)\} \\ \text{s.t.} \ & \phi_0\sigma(g|G) \geq (1-\phi_0)\sigma(g|N) \\ & (1-\phi_0)\sigma(n|N) \geq \phi_0\sigma(n|G) \\ & \sigma(g|G) + \sigma(n|G) = 1 \\ & \sigma(g|N) + \sigma(n|N) = 1 \end{aligned}$$

  (the latter two are the feasibility constraints on σ)

- Solution: $\sigma(g|G) = 1$, $\sigma(n|N) = \frac{1-2\phi_0}{1-\phi_0}$. Again, same as we had.

Conclusions

- We have seen two approaches to information design (equivalent when both are applicable).
- Insight from example:
  - make player take undesirable action only if completely confident it is the right one;
  - make player barely indifferent when taking the desirable action (because then you can convince them to take it more frequently).
- What we have not seen:
  - With many players, sometimes optimal to correlate messages positively (i.e. public experiments optimal), sometimes negatively – depends on the game.
  - Dynamic problems are also interesting (how to optimally reveal over time the information that arrives dynamically).
- Hot field but it is not that easy to find a real-life setting to which it could apply, because of designer's commitment.

References I

D. Bergemann and S. Morris. Information Design: A Unified Perspective. *Journal of Economic Literature*, 57(1):44–95, Mar. 2019. ISSN 0022-0515. doi: 10.1257/jel.20181489. URL `https://pubs.aeaweb.org/doi/10.1257/jel.20181489`.

J. Best and D. Quigley. Persuasion for the Long Run. 2020. URL `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2908115`.

E. Bondi, H. Oh, H. Xu, F. Fang, B. Dilkina, and M. Tambe. To Signal or Not To Signal: Exploiting Uncertain Real-Time Information in Signaling Games for Security and Sustainability. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(02): 1369–1377, Apr. 2020. ISSN 2374-3468, 2159-5399. doi: 10.1609/aaai.v34i02.5493. URL `https://ojs.aaai.org/index.php/AAAI/article/view/5493`.

References II

E. Kamenica. Bayesian Persuasion and Information Design. *Annual Review of Economics*, 11 (1):249–272, Aug. 2019. ISSN 1941-1383, 1941-1391. doi: 10.1146/annurev-economics-080218-025739. URL `https://www.annualreviews.org/doi/10.1146/annurev-economics-080218-025739`.