Attention Capture

Andrew Koh Sivakorn Sanguanmoo (Win)
MIT Economics

2025 ASSA Annual Meeting

Some motivation

- Our consumption of information is (i) dynamic; and (ii) channeled through a designer/algorithm:
  - Search engines, social media, streaming platforms etc.
- These platforms have incentive to keep us on them:
  - 2022 Q1: 97% of Facebook's revenue, 81% of Google's revenue, and 92% of Twitter's from ads

BUSINESS • STREAMING

How Ads on Netflix Will Change the Way You Watch

Apple Finds Its Next Big Business: Showing Ads on Your iPhone

Ad infinitum: companies to unleash a deluge of digital marketing

Delivery apps, ecommerce marketplaces, mass market retailers, gaming services all target commercials for revenue

Instagram to increase ad load as Meta fights revenue decline

- This paper: what are the limits of information to capture attention? How much commitment is required?

Outline

Setting

- Single decision marker with preferences over (actions, states, time)
- Fix a dynamic info structure (for each state, time, history of messages, specifies distribution of message) $\rightarrow$ DM stops & acts at some random time.

Questions

1. How is attention optimally extracted?
   - We solve this using reduction principle
   - Characterize convex-order frontier and extreme points
2. How does equilibrium change if designer has commitment vs not?
   - No commitment gap: for arbitrary DM & designer preferences, optimal structures have sequentially optimal modifications
3. How do we optimally extract attention & persuade?
   - We solve this for binary states/actions [Not covered today]

(Brief) Literature

1. **Dynamic info design where info valuable for action.**
   - Knoepfle (2020); Hébert and Zhong (2022)
   - Our work: nonlinear designer's value
   - Saeedi et al. (2024): similar baseline model but different approaches and behavioral extensions
2. **Dynamic info design where info valuable for stopping.**
   - Ely and Szydlowski (2020); Orlov et al. (2020)
   - We show that no commitment is necessary in general.
3. **Info acquisition:** DM in control of info structure. Zhong (2022)
   - Also: Pomatto et al. (2018), Morris and Strack (2019) etc.
4. **Sequential learning/sampling.** Starting from Wald (1947) and Arrow, Blackwell, and Girshick (1949).

Model 1/2

- Finite states Θ, actions A, time discrete $\mathcal{T} = 0, 1, \ldots$
- DM has full-support prior $\mu_0 \in \Delta(\Theta)$ and has payoff function $v : A \times \Theta \times \mathcal{T}$ from taking action a under state θ at time τ :

$$v(a, \theta, \tau) := u(a, \theta) - c\tau.$$

- $I \in \Delta(\prod_{t \quad 1} \Delta(\Theta))$ is a dynamic info structure if for any μ_t and H_t,

$$\mu_t = \int_{\mu_{t+1}, m} \mu_{t+1} dI_{t+1}(\mu_{t+1} | H_t)$$

  $I_{t+1}(\cdot | H_t)$ is cond. dist. over next period's belief
- DM solves

$$\sup_{\tau, a_\tau} \mathrm{E}^I [v(a_\tau, \theta, \tau)]$$

  E^I is expectation under I, and (τ, a_τ) are w.r.t. natural filtration. Assume tiebreak to not stop. $\mathcal{I}$ is set of all dynamic info.

Model 2/2

- DM's optimal stopping gives map $I \mapsto d(I) \in \Delta(\mathcal{T})$.
- $d \in \Delta(\mathcal{T})$ is **feasible** if there exists info structure I such that $d = d(I)$.
- Designer has preferences $f : \mathcal{T} \to \mathbb{R}$. With commitment, solves

$$\sup_{I \in \mathcal{I}} \mathrm{E}^I[f(\tau)]$$

- Implicit assumptions
  - Full commitment: no need for intertemporal commitment
  - Pure attention capture: platform primarily aims to extract attention not persuasion. Add persuasion aspect in paper

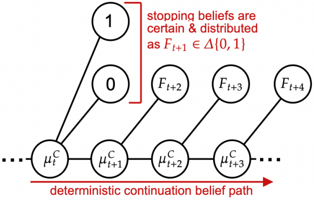
Reduction Principle

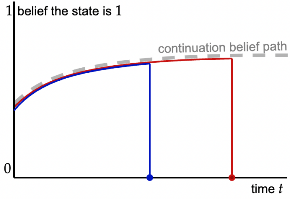
- The space of info structures is large $\rightarrow$ need to narrow down

Definition

I is full-revelation with deterministic continuation beliefs if there exists a unique belief path $(\mu_t^C)_t$ such that for any H_t with prob > 0

1. (Full revelation) supp $I_{t+1}(\cdot \mid H_t) \subset \underbrace{\{\mu_{t+1}^C\}}_{\text{continue}} \cup \underbrace{\{\delta_\theta : \theta \in \Theta\}}_{\text{full info + stop}}$
2. (Obedience) For each t, DM prefers to continue at history $H_t = (\mu_s^C)_{s \quad t}$ and stop at $H_t = (\mu_0, \mu_1^C, \ldots, \mu_{t \quad 1}^C, \delta_\theta)$.





Optimal attention capture: reduction

Proposition (Reduction principle for attention)

If $d \in \Delta(\mathcal{T})$ is feasible it can be implemented by some full-revelation & obedient structure

- Quite useful for optimization, intuition related to revelation principle.
- Whenever DM stops, give her full info - $\uparrow$ info value $\Rightarrow$ no change in stopping time as continuation incentives are preserved
- Collapse all continuation nodes into a single node "continue"

Writing down obedience constraints explicitly

- Recall: $(\mu_t^C)_t \in \prod_{t \geq 1} \Delta(\Theta)$ is a belief path associated with full-revelation and obedient structure I
- Value of full info under belief μ :

$$\phi(\mu) := \mathsf{E}_\mu[\max_{a \in A} u(a,\theta)] - \max_{a \in A} \mathsf{E}_\mu[u(a,\theta)]$$

"At belief μ, what's my value of learning the state vs acting now?"

- Obedience at time-t requires

$$\underbrace{\phi(\mu_t^C) \geq \overbrace{\mathsf{E}[c\tau \mid \tau > t] - ct}^{\text{attention cost until stop}}}_{\textbf{'Obedience constraint'}}$$

$$\boxed{\Phi \;:= \text{argmax}_{\mu \in \Delta(\Theta)} \phi(\mu) \subseteq \Delta(\Theta)} \qquad \boxed{\phi \;:= \max_{\mu \in \Delta(\Theta)} \phi(\mu)}$$

Φ $=$ Basin of uncertainty (beliefs that have the highest value of full info)

Full-rev. & Obedient $\leftrightarrow$ Belief Path & Stopping Time

- So far obedience constraint: continuing is better than stopping.
- Not the only constraint: fixing τ, we're not free to pick any continuation belief.
- Boundary constraint: For every $t \in \mathcal{T}$ and $\theta \in \Theta$,

$$\mathsf{P}^I(\tau > t+1)\mu_{t+1}(\theta) \leq \mathsf{P}^I(\tau > t)\mu_t(\theta).$$

- Idea: Apply the martingale property of beliefs given $\tau > t$:

$$\mu_t(\theta) = 1 \cdot \mathsf{P}^I(\mu_{t+1} = \delta_\theta \mid \tau > t) + \mu_{t+1}(\theta) \cdot \underbrace{\mathsf{P}^I(\tau > t+1 \mid \tau > t)}_{\text{Prob. don't get full info}}$$

$$\geq \mu_{t+1}(\theta)\mathsf{P}^I(\tau > t+1 \mid \tau > t)$$

  - Clearly necessary, but boundary + obedience also sufficient!

Lemma

The following are equivalent:

1. *There exists a full-revelation and obedient information structure* $I \in \mathcal{I}^{FULL}$ *which induces stopping time* $\tau(I)$ *and belief path* $(\mu_t^C)_{t2T}$.
2. *The following conditions are fulfilled:*
   (i) *(Obedience constraint)* $\phi(\mu_t^C) \geq \mathrm{E}[c\tau \mid \tau > t] - ct$ *for every* $t \in \mathcal{T}$; *and*
   (ii) *(Boundary constraint)* $\mathrm{P}^I(\tau > t+1)\mu_{t+1}^C(\theta) \leq \mathrm{P}^I(\tau > t)\mu_t^C(\theta)$ *for every* $t \in \mathcal{T}$ *and* $\theta \in \Theta$.

- Reduced our problem to finding pair of belief paths and stopping time which satisfies obedience and boundary:

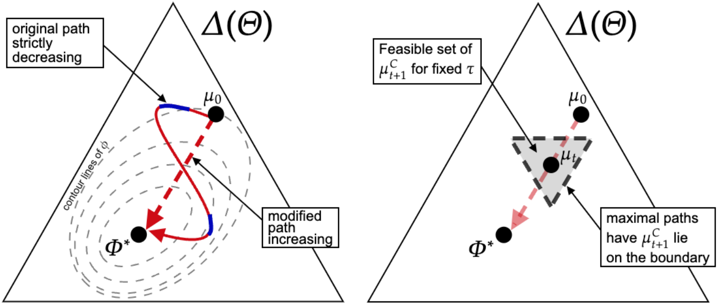
$$f_{\mu_0} := \max_{\substack{(d_T(\tau),(\mu_t^C)_t) \\ 2\Delta(T) \quad (\Delta(\Theta))^T}} \mathrm{E}^I[h(\tau)] \qquad \boxed{\text{Original program}}$$

$$\text{s.t.} \quad \phi(\mu_t^C) \geq \mathrm{E}[c\tau \mid \tau > t] - ct \quad \forall t \in \mathcal{T} \qquad \text{(Obedience)}$$

$$\mathrm{P}(\tau > t+1)\mu_{t+1}^C \leq \mathrm{P}(\tau > t)\mu_t^C \qquad \text{(Boundary)}$$

Increasing and Maximal Belief Paths

- Belief path $(\mu_t^C)_t$ is **increasing** if $(\phi(\mu_t^C))_t$ is increasing.
- Belief path $(\mu_t^C)_t$ is **maximal** for stopping time τ if Boundary constraints bind whenever $\mu_{t+1}^C \notin \Phi$, i.e., μ_{t+1}^C has not reached basin of uncertainty Φ yet.



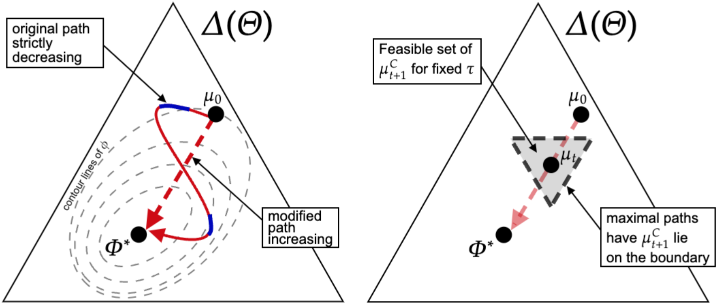
(a) Increasing paths (b) Maximal paths

Increasing and maximal are sufficient

Theorem

Every feasible stopping time can be implemented through full-revelation and deterministic structures with increasing and maximal continuation belief paths.

- Sufficient to consider belief path **maximally** steering toward basin of uncertainty $\rightarrow$ smaller space to consider to solve designer's optimum



Optimal attention capture: concave value

Suppose h is concave. Obedience at time 0 implies

$$\mathsf{E}[c\tau] \leq \phi(\mu_0).$$

By Jensen's inequality, need to concentrate stopping time:

$$\mathsf{E}[h(\tau)] \leq h(\phi(\mu_0)/c).$$

Proposition

Suppose $\phi(\mu_0)/c$ is integer. the optimal info structure under concave to reveal full info at time $T = \phi(\mu_0)/c$ and $\tau = \phi(\mu_0)/c$ a.s.

Optimal attention capture: convex value

Suppose h is convex.

- Obedience at time 0 gives upper bound of average stopping time $E[\tau] \leq \phi(\mu_0)/c$.

Designer wants to spread stopping time as much as possible.

- Main concern: obedience constraints must hold for all times
- "Give info at time 0; otherwise, give info at very large time" violates obedience condition since DM stops paying attention if she gets no info at time 0.

Our approach: characterize convex order frontier

- Recall: d dominates d^0 in convex order, i.e., $d \succeq_{cx} d^0$ if $E_{\tau \quad d}[f(\tau)] \geq E_{\tau \quad d^0}[f(\tau)]$ for any convex function $f : \mathcal{T} \to R$.

IIM distribution

Definition (Indifference, increasing, and maximal (IIM) distribution)

$d \in \Delta(\mathcal{T})$ is an indifference, increasing, and maximal (IIM) distribution if

1. $\exists \boldsymbol{\mu}^C$ s.t. $(d, \boldsymbol{\mu}^C)$ is feasible, $\boldsymbol{\mu}^C$ increasing and maximal + Obedience binds for all $t \geq 1$
2. $(d, \boldsymbol{\mu}^C)$ feasible $\Rightarrow \boldsymbol{\mu}^C$ increasing and maximal.

- Obedience binds for all t : DM is indifferent between continuing and stopping every period.
  - Common in literature but not sufficient to pin down structure
- + Increasing and maximal belief path
  - Help pin down optimal info structure especially binary states
  - This property is also a necessity condition.

Convex-order frontier

Theorem

For any feasible stopping time d, there exists an indifferent, increasing, and maximal distribution d^{IIM} for which

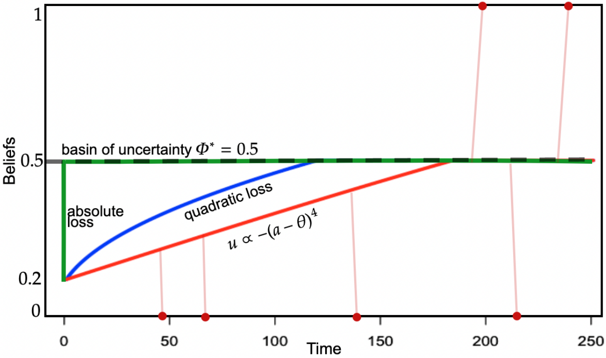
$$d^{IIM} \succeq_{CX} d.$$

This implies if d is not IIM then it is not on the convex-order frontier i.e., the relation is strict.

- Best (and necessary) way to spread stopping time is
  1. to make DM indifferent at every time (so that DM pays attention in longer period) while
  2. to steer DM's continuation belief toward the basin of uncertainty Φ as much as possible

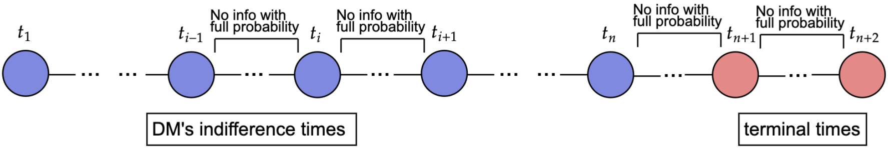
Convex-order frontier: optimal belief paths

- Recall obedience constraint: $\phi(\mu_t^C) \geq \mathsf{E}[c\tau \mid \tau > t] - ct$
- For convex frontier, it is necessary to have a wide range of stopping time
  - Steering DM's continuation belief Φ is necessary so that value of full info becomes higher over time.
- When $|\Theta| = 2$, belief path that binds Obedience every time is uniquely pinned down by increasing and maximal conditions.



Exotic designer's preferences (If time permits)

- Designer's preference might be neither concave nor convex
  - S-shaped function: users are highly responsive to advertising at some intermediate times
- Characterize extreme points of feasible stopping times for binary actions and states: each extreme point is induced by a "block structure"
  - A "block" is a time period between two adjacent times in support.
  - Block structure: DM is indiff at a starting time of every block (except the last)
- Support of stopping time pins down block structure because of indifference + increasing and maximal belief path
  - In paper, apply block structure to solve attention capture under S-shaped function



Time-consistency

- So far: Designer can commit future info structures $\rightarrow$ intertemporal commitment.
- How do results change when no intertemporal commitment power?

Definition

I is **sequentially optimal** for designer preference f if, for every history H_t with positive probability,

$$\max_{I' \in \mathcal{I}|H_t} \mathrm{E}^{I'}\left[f\left(\tau(I')\right)\middle|H_t\right] = \mathrm{E}^{I}\left[f\left(\tau(I)\right)\middle|H_t\right]$$

where $\mathcal{I}|H_t$ is the set of info structures where H_t realizes with positive probability.

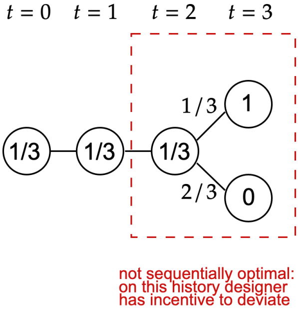
- At every history, designer has no incentive to different continuation info structure.
- If I is sequentially optimal, I is also optimal.
    - Existence of sequentially optimal info structure $\rightarrow$ no need for intertemporal commitment.

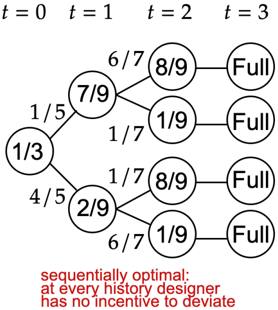
An intuitive example:

- $A = \Theta = \{0, 1\}$
  $v(a, \theta, t) = -(a - \theta)^2 - ct \leftarrow$ waiting costly, constant per-unit
  $c = 1/9$, $\mu_0 := \mathrm{P}(\theta = 1) = 1/3$.
- $f(a, \tau) = \tau \leftarrow$ linear value of attention
- The DM's payoff from stopping and taking action at time $t = 0$ is $-\frac{1}{3}$.
- Obedience at time 0:

$$-\mathrm{E}[c\tau] \geq -1/3 \Rightarrow \mathrm{E}[\tau] \leq (1/c) \cdot (1/3) = 3$$

An intuitive example: optimal info





LHS: Optimal but not sequentially optimal

- Conditional on the DM continues until $t = 2$, designer can deviate to reveal full info at $t = 4 \Rightarrow$ DM still wants to follow.

RHS: Optimal & sequentially optimal

- Conditional on the DM continues until $t = 2$, designer cannot delay full info to $t = 4$ because optimal util under belief $8/9$ is $-1/9 = -c$.

No intertemporal gap

Theorem

For arbitrary DM's and designer's util functions, sequentially optimal dynamic info structures exist.

- Every optimal info structure can be modified so that it is also sequentially optimal.
  - Info must be gradually delivered
  - No longer deterministic continuation beliefs

Proof Sketch

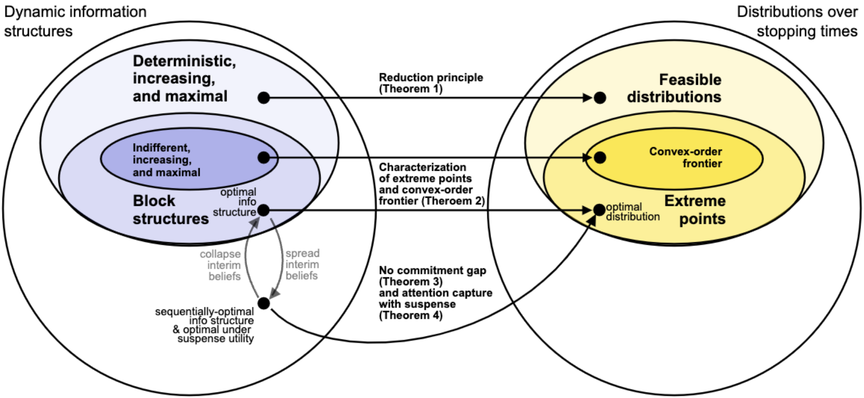
- Key step: If I is optimal and DM is indiff between continuing and stopping at every history, then I is also sequentially optimal.
- Perform surgery on optimal info structure so that DM is indiff at every history.
  - Anti-deterministic: spreading continuation beliefs

Our subsequent work (Koh et al., 2024) generalizes no-commitment gap result to arbitrary dynamic info design with optimal stopping.

Concluding remarks

- Solve optimal attention capture and show no intertemporal commitment gap
- Not covered today: Noninstrumental value of info and attention capture with persuasion motives

Figure: Connections between aspects of attention capture



References

Arrow, K. J., D. Blackwell, and M. A. Girshick (1949): "Bayes and minimax solutions of sequential decision problems," *Econometrica, Journal of the Econometric Society*, 213–244.

Ely, J. C. and M. Szydlowski (2020): "Moving the goalposts," *Journal of Political Economy*, 128, 468–506.

Hébert, B. and W. Zhong (2022): "Engagement Maximization," *arXiv preprint arXiv:2207.00685*.

Knoepfle, J. (2020): "Dynamic Competition for Attention," Tech. rep., University of Bonn and University of Mannheim, Germany.

Koh, A., S. Sanguanmoo, and W. Zhong (2024): "Persuasion and Optimal Stopping," *arXiv preprint arXiv:2406.12278*.

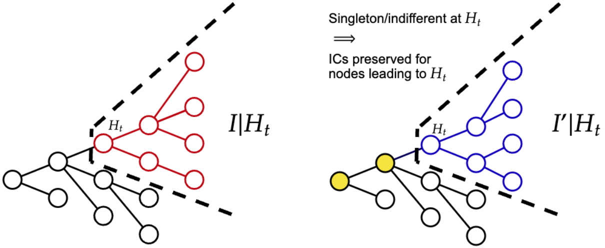
Morris, S. and P. Strack (2019): "The wald problem and the relation of sequential sampling and ex-ante information costs," *Available at SSRN 2991567*.

Orlov, D., A. Skrzypacz, and P. Zryumov (2020): "Persuading the principal to wait," *Journal of Political Economy*, 128, 2542–2578.

Pomatto, L., P. Strack, and O. Tamuz (2018): "The cost of information," *arXiv preprint arXiv:1812.04211*.

Optimal + indiff at each time $\Longrightarrow$ sequentially optimal

- Let I be opt and DM is indiff for each time, suppose not seq. opt at H_t
- Designer can strictly do better by changing $I|H_t$ to $I^0|H_t$
- If this preserves DM's stopping/continuing IC at earlier times t, then this contradicts the optimality of I!
  - For $s \leq t$ and connected to H_t, was previously continuation at I, still want to continue $\leftarrow$ we need to show this!
  - Everything else remains the same:



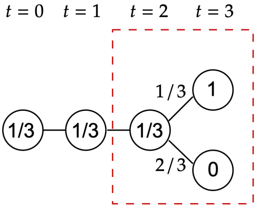
Implies overall strictly better for the designer (why?)

Still need to show continuation incentive at H_t increases

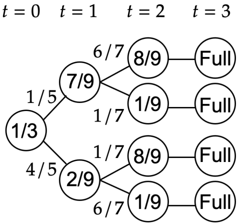
- Let $V^I(H_t) := \sup_{\tau, a_\tau} \mathrm{E}^I[v|H_t]$ $\boxed{\text{WTS } V^{I^0}(H_t) \geq V^I(H_t)}$
- Since DM is indifferent,

$$V^I(H_t) = \max_{a \in A} \mathrm{E}[v(a, \theta, t)] \leq V^{I^0}(H_t)$$

Key intuition: outside option of stopping & acting is a lower bound on DM's continuation payoff



not sequentially optimal: on this history designer has incentive to deviate



sequentially optimal: at every history designer has no incentive to deviate