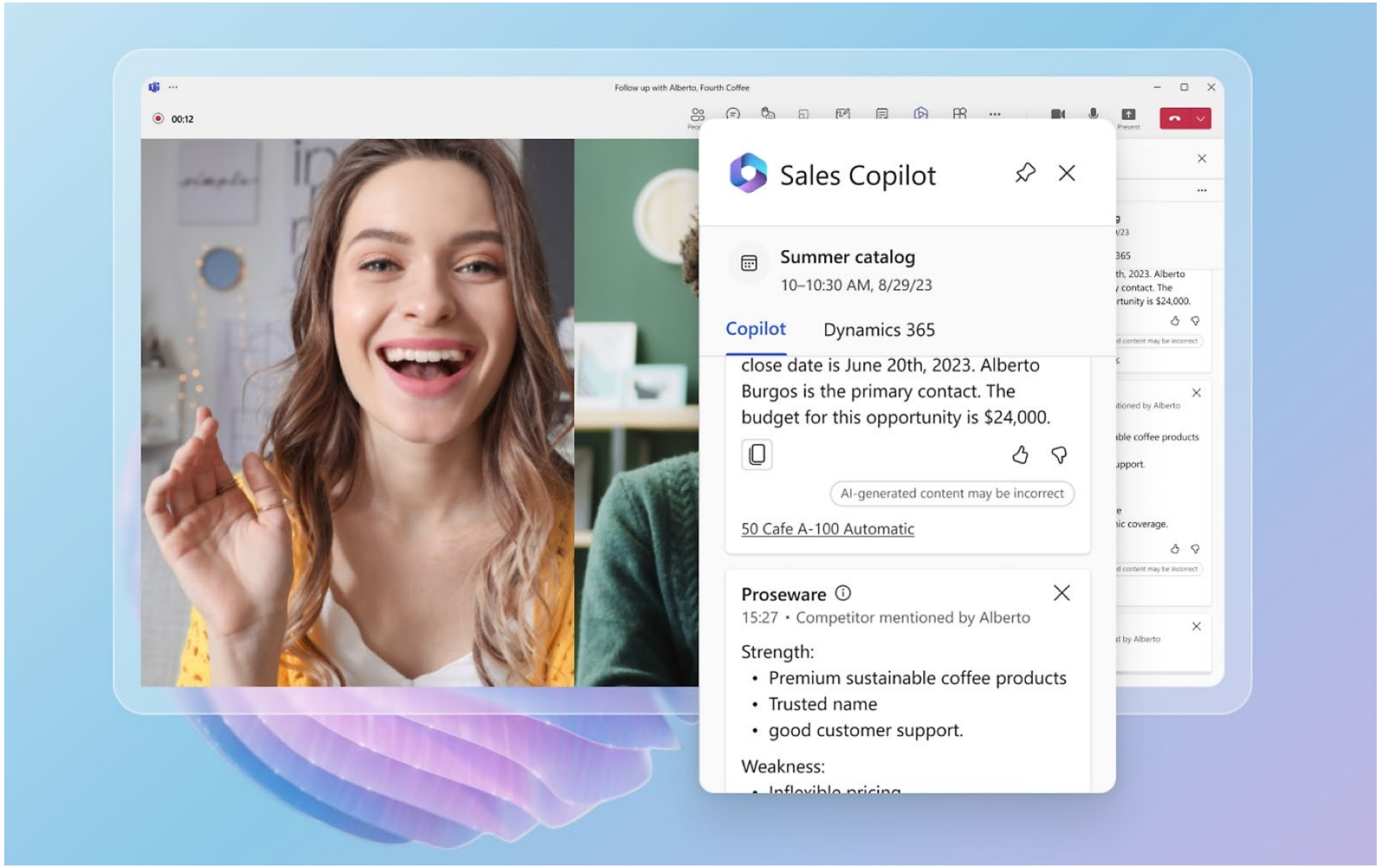
Algorithmic Persuasion through Simulation

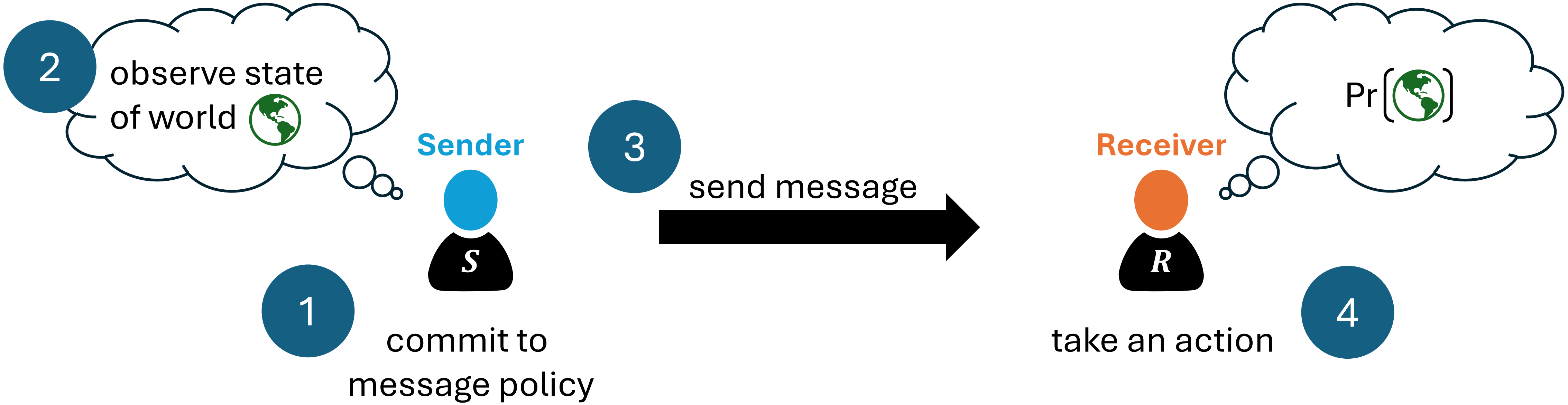
Nicole Immorlica, Microsoft Research

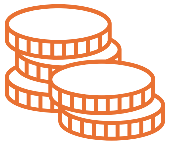
Based on joint work with Keegan Harris,
Brendan Lucier, and Alex Slivkins

AI and persuasion.



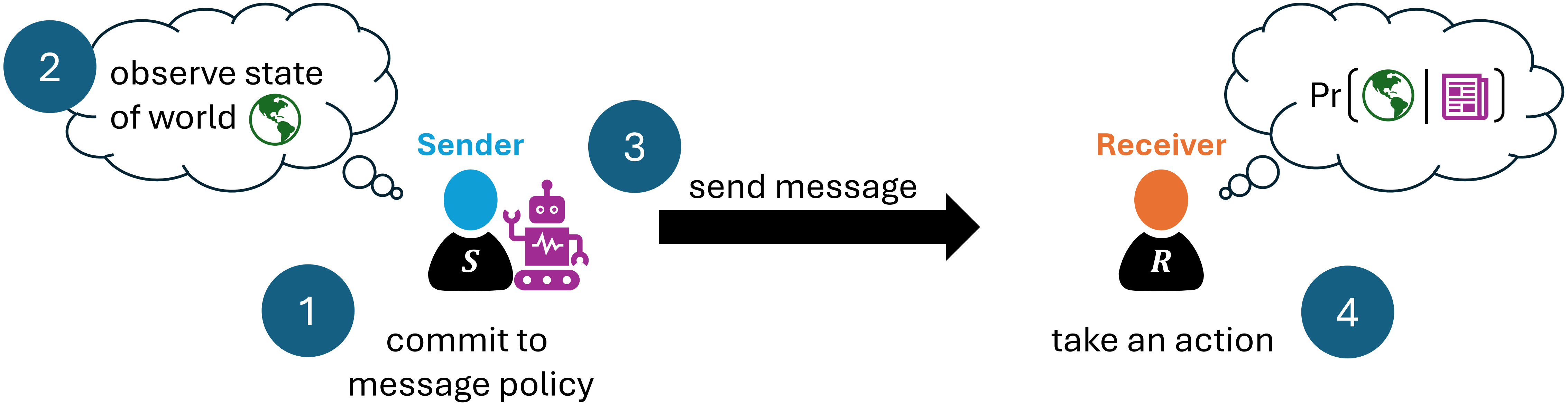
persuasion.





Utilities are function of state and action.

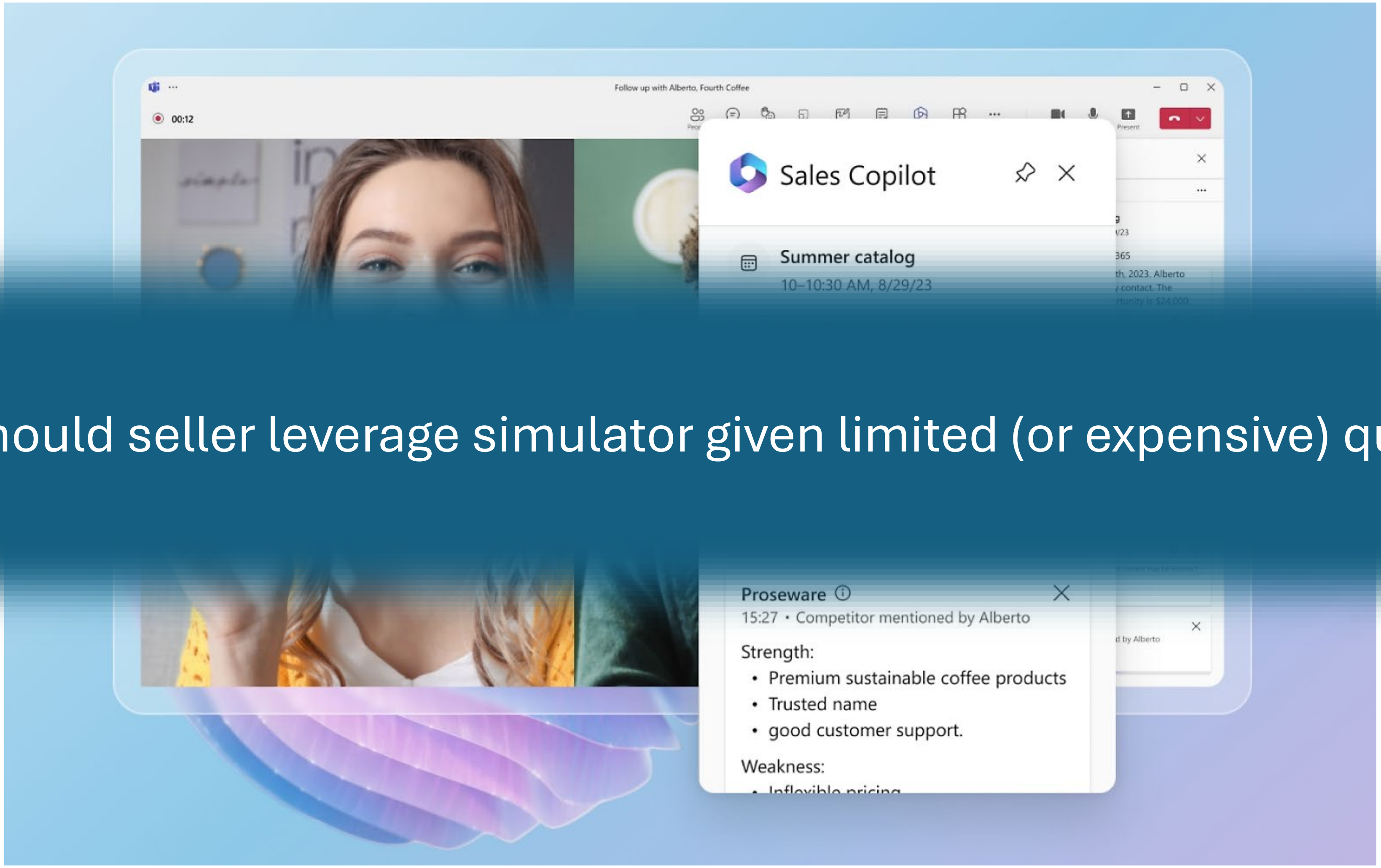
persuasion.



Receiver observes signal correlated with state.

Simulation oracle tells sender how receiver will react.

sales copilot.



How should seller leverage simulator given limited (or expensive) queries?

related work.

Bayesian persuasion with informed receivers:

- Optimal experiment is a linear (or convex) optimization problem [Gentzkow and Kamenica 2016], [Candogan 2019], [Candogan and Strack 2022]
- Screening is equivalent to experiments with binary actions, but not otherwise [Kolotilin et al. 2017], [Guo and Shmaya 2019], [Candogan and Strack 2022]

Pure exploration in bandits and learning in Stackelberg games:

- Predict best action after K rounds of exploration [Bubeck et al. 2009], [Chen et al. 2014], [Xu et al. 2018]
- Learn optimal strategy for leader in Stackelberg games from query access [Letchford et al. 2009], [Balcan et al. 2015], [Peng et al. 2019]

model.

State: $\omega \in \{0,1\}$ representing quality of product (high or low)

Receiver/buyer: binary action $a \in \{0,1\}$ representing purchasing decision

- has private signal τ drawn from finite set T*
- private signal is correlated with state, i.e., $(\omega, \tau) \sim F$
- utility = $\begin{cases} 1 \text{ if purchased product and high quality} \\ -1 \text{ if purchased and low quality} \\ 0 \text{ otherwise} \end{cases}$

* Can also handle continuum signal space via discretization

Sender/seller:

- can commit to policy $\sigma: \{0,1\} \rightarrow M$ mapping state to messages
- utility 1 if product purchased, 0 otherwise

simulation oracle.

A black-box that simulates receiver's action for any message.

Definition: A simulation oracle inputs a query consisting of a messaging policy σ and a message m and returns receiver's optimal action given posterior beliefs, i.e.,

$$\text{argmax}_a \, E_\omega[u_R(\omega, a) | \sigma(\omega) = m, \tau].$$

Examples: generative AI (e.g., sales copilot), survey/historical data, sequence of myopic buyers, algorithmic buyer agents (e.g., autobidders in ad auctions)

game.

Timing:

1. State $\omega \in \{0,1\}$ and receiver's private signal $\tau \in T$ drawn from joint distribution F
2. Sender adaptively queries simulator up to K times
3. Sender commits to random message policy σ mapping states to messages $m \in M$
4. Sender observes state and sends signal $m \sim \sigma(\omega)$ to receiver
5. Receiver takes action $a \in \{0,1\}$

Definition: A query policy π maps a history $h \in H$ of queries and responses to a new query (overloading notation, let $\pi(\tau) \in H$ be history generated by π when signal is τ)

equilibrium.

Perfect Bayesian Equilibrium.

- Receiver takes utility-maximizing action given belief induced by signal and message.
- Sender chooses utility-maximizing messaging policy and query policy given that receiver behaves in this manner.

Definition: Strategies (π^*, σ^*, a^*), belief $B_S: H \to \Delta(T)$ for the sender mapping query histories to distributions over receiver signals, and belief $B_R: M \times T \to \Delta(\{0,1\})$ for the receiver mapping messages and signals to distributions over the state, is a PBE if:

- For each m and τ, action $a^*(m,\tau)$ maximizes receiver's utility given belief $B_R(m,\tau)$
- Belief $B_R(m,\tau)$ is posterior distribution over ω given τ, σ^*, and fact that $\sigma^*(\omega) = m$
- For each $h \in H$, message policy σ^* maximizes sender's utility given belief $B_S(h)$
- Belief $B_S(h)$ is posterior distribution over τ, given π^* and fact that $\pi^*(\tau) = h$
- Sender's querying policy π^* maximizes sender's utility given σ^* and a^*

results.

Theorem: Can compute sender-optimal PBE strategies (π^*, σ^*, a^*) in time polynomial in number of private signals $|T|$.

Proof overview: Show how to compute in polynomial time,

1. Receiver action a^* given message and message policy
2. Sender message policy σ^* given query policy π^* and query responses $\pi^*(\tau)$
3. Sender query policy π^*

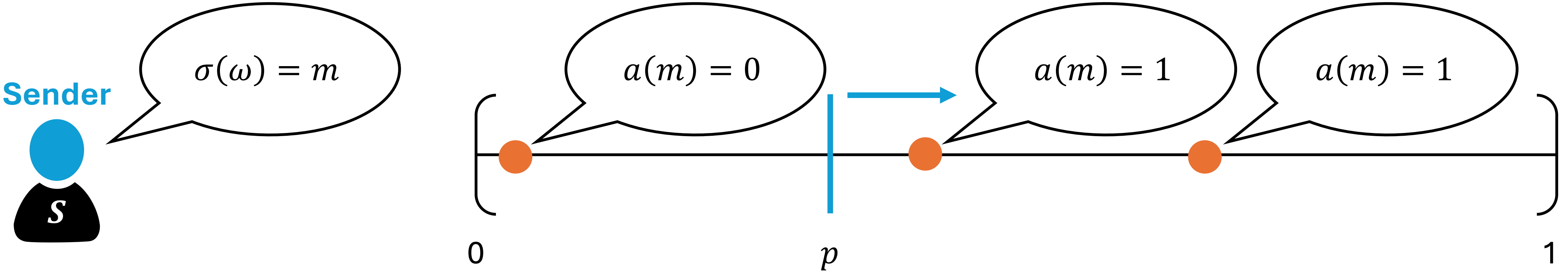
Implications and extensions:

- Structure: optimal query policy precomputes a pooling of receivers into contiguous intervals of beliefs, then uses queries to identify interval and message policy
- Robustness to noise: message policy robust to slight downward perturbations of beliefs (i.e., assuming receiver is slightly more pessimistic than queries suggest)
- Myopic receivers: query policy trades off between exploration and exploitation

optimal message policy.

Proposition: For any messaging policy σ, there is an outcome-equivalent messaging policy σ' with just $|M| = |T| + 1$ messages.

Proof Sketch: Consider any message $m \in M$



(abusing terminology, will equate signal with induced posterior belief or type)

optimal message policy.

Proposition: For any messaging policy σ, there is an outcome-equivalent messaging policy σ' with just $|M| = |T| + 1$ messages.

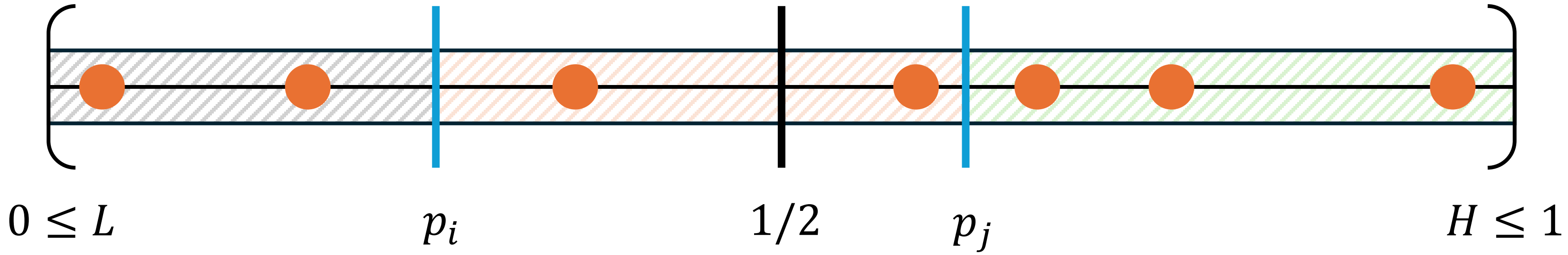
Implication: There is a 1:1 correspondence between messages and thresholds.

- sender can distinguish between any pair of beliefs with one simulation query
- can uniquely identify receiver belief with $\log |T|$ simulation queries
- optimal querying policy for $K < \log |T|$ equates to choosing set of thresholds

Note: In sufficiently rich economic environments, a single query may suffice to identify belief.

optimal message policy.

Proposition: For any given set T of beliefs, the optimal messaging policy mixes between a two messages m_i and m_j signifying threshold beliefs.



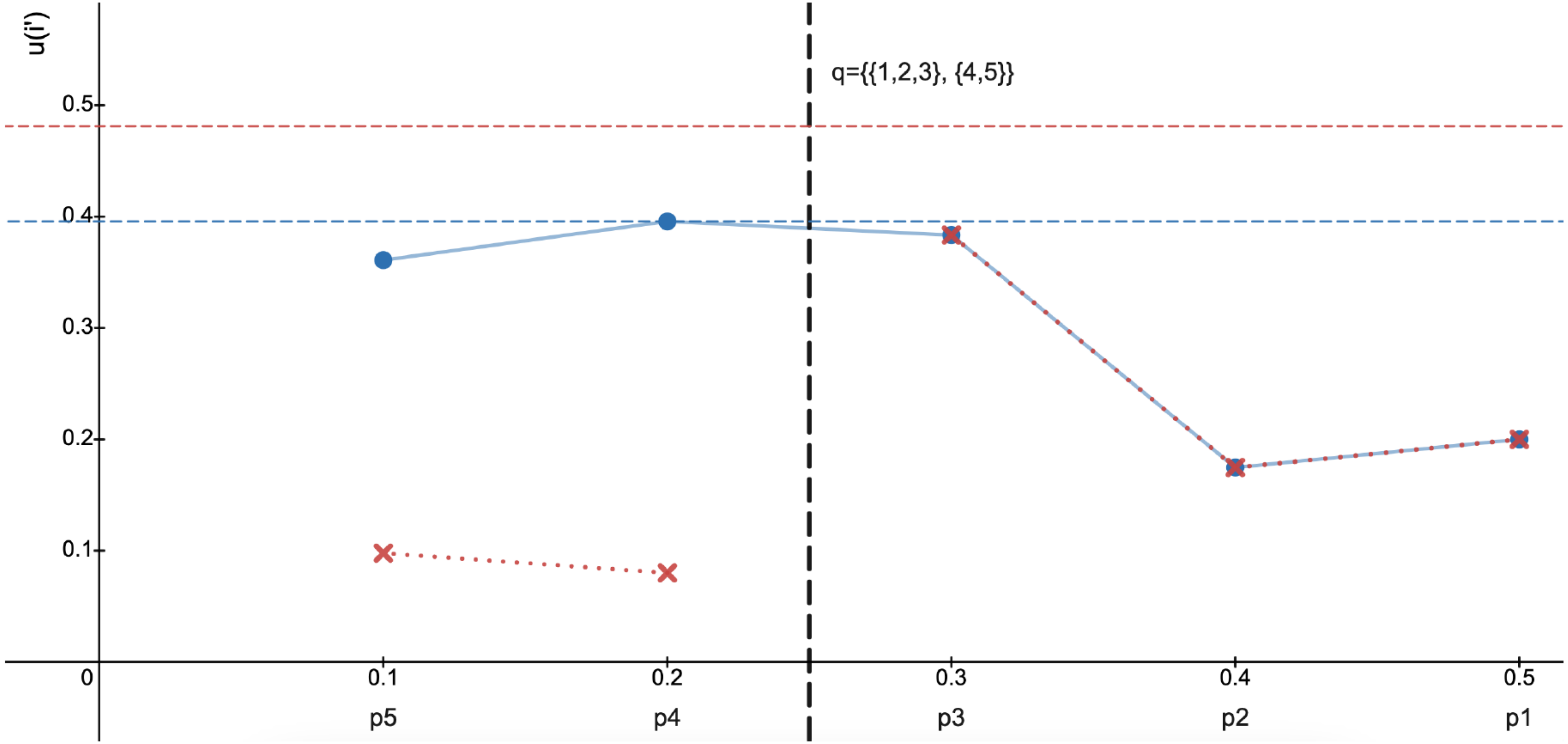
Proof Sketch: Given the correspondence between messages and thresholds,

- optimization problem is a linear program with $|T| + 2$ constraints
- substituting for tight IC constraints leaves just two non-trivial constraints
- rank lemma implies optimal solution has positive weight on just two variables

Note: The policy can be computed in time $O(|T|^2)$.

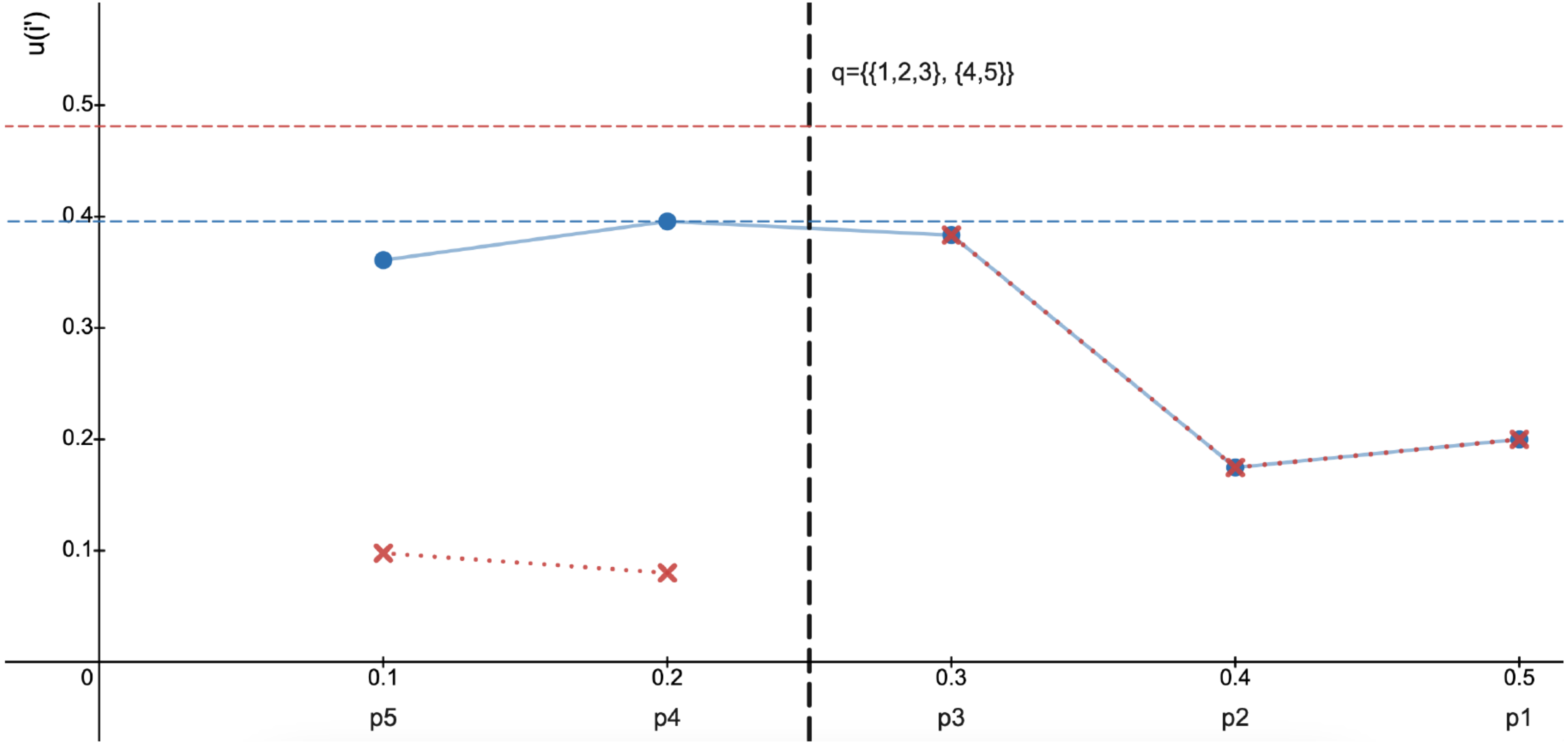
optimal message policy.

Instance: Posterior beliefs $(0.1, 0.2, 0.3, 0.4, 0.5)$ w/prob $(0.2, 0.2, 0.39, 0.01, 0.2)$, resp.



optimal query policy.

Gain from single query: submodular, implying greedy is constant-factor approximation



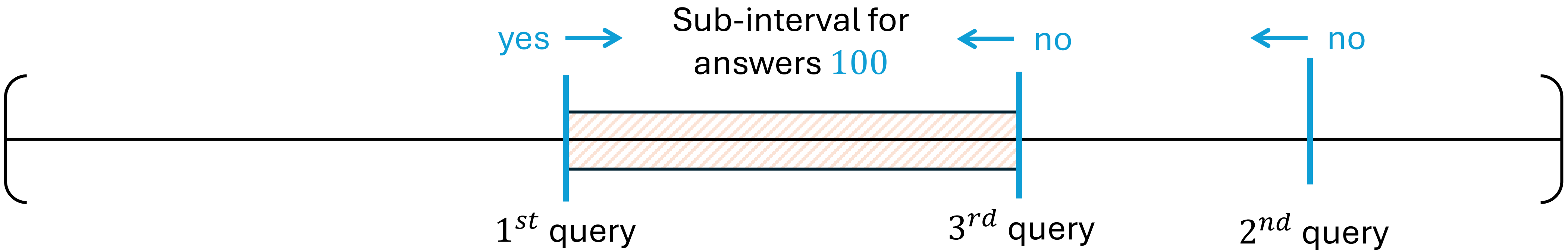
optimal query policy.

Theorem: The optimal query policy can be computed in time $O(|T|^2 \min\{|T|, 2^K\})$ where K is the bound on the number of queries.

Lemma: Suffices to compute a set of $\min\{|T|, 2^K\}$ possible queries and adaptively choose among them using binary search.

Proof: equivalence between K-query adaptive and 2^K-query non-adaptive policies

1. A given adaptive policy partitions space into $\leq 2^K$ sub-intervals of beliefs.



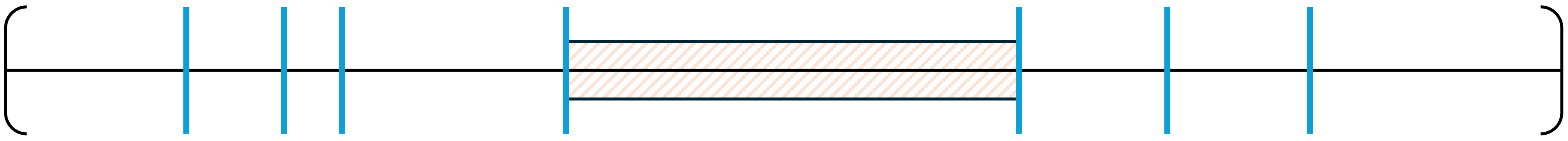
optimal query policy.

Theorem: The optimal query policy can be computed in time $O(|T|^2 \min\{|T|, 2^K\})$ where K is the bound on the number of queries.

Lemma: Suffices to compute a set of $\min\{|T|, 2^K\}$ possible queries and adaptively choose among them using binary search.

Proof: equivalence between K-query adaptive and 2^K-query non-adaptive policies

2. A non-adaptive policy that queries the $2^K - 1$ thresholds separating the sub-intervals of an adaptive policy gains same information resulting in same value.



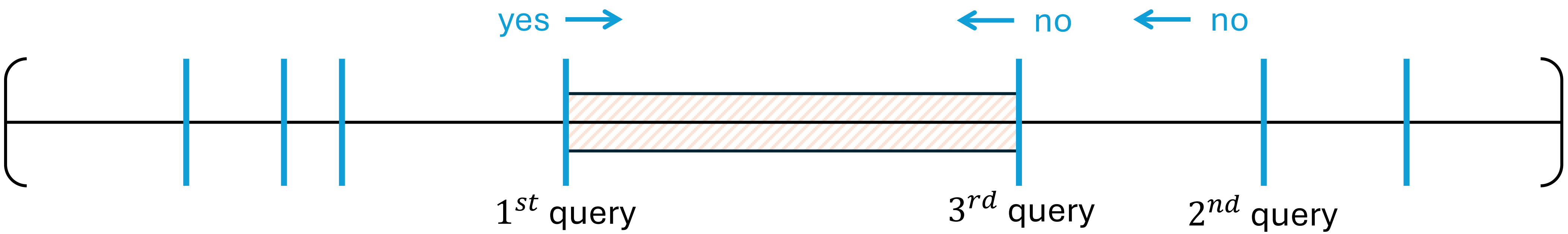
optimal query policy.

Theorem: The optimal query policy can be computed in time $O(|T|^2 \min\{|T|, 2^K\})$ where K is the bound on the number of queries.

Lemma: Suffices to compute a set of $\min\{|T|, 2^K\}$ possible queries and adaptively choose among them using binary search.

Proof: equivalence between K-query adaptive and 2^K-query non-adaptive policies

3. An adaptive policy performs binary search among the $2^K - 1$ queries of a non-adaptive policy gains the same information and hence the same value.

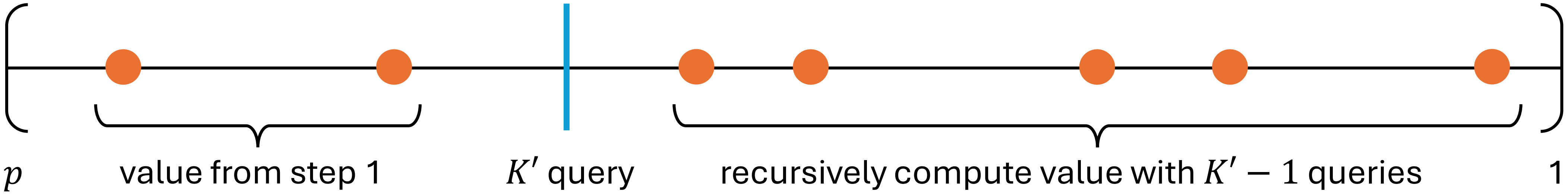


optimal query policy.

What set of queries should sender select to maximize utility?

Dynamic Program: Compute optimal set of 2^K non-adaptive queries.

1. Compute optimal message policy for each of the $|T|^2$ subinterval of types.
2. Optimal value of K' queries for subinterval of types from p to 1 is sum of best split given $K' - 1$ remaining queries in suffix.

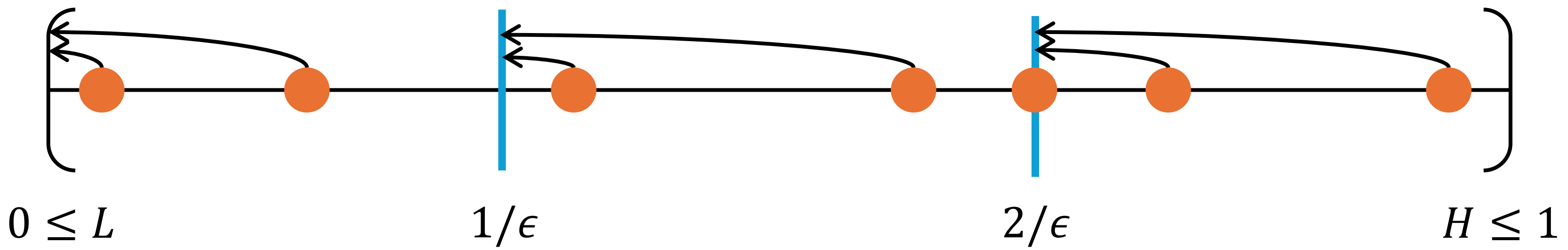


3. Return queries computed from using 2^K queries for interval of all types.

optimal query policy.

Theorem: The optimal query policy can be computed in time $O(|T|^2 \min\{|T|, 2^K\})$ where K is the bound on the number of queries.

Note: Value of policy is robust to perturbations of thresholds, so at an additive loss of ϵ to sender's utility, can run in time $O(\epsilon^{-2} \min\{1/\epsilon, 2^K\})$.



OPT-Rounded(original) ≥ OPT-Rounded(rounded) ≥ OPT-Original(original) - ϵ

generalizations.

Approximate oracles: If difference between true and assumed belief is at most δ with probability at least $1-\gamma$, policy guarantees at least $(1-\gamma)OPT - O(\delta)$.

Costly queries: Suppose each query q has a cost c_q to the sender. Then the following dynamic program $A(1,|T|)$ computes the optimal query policy in time $O(|T|^3)$.

$$A(1,j) \coloneqq \max\left\{V[1,j];\max_{q\in T} A[1,q]+V[q+1,j]-c_q\right\}$$

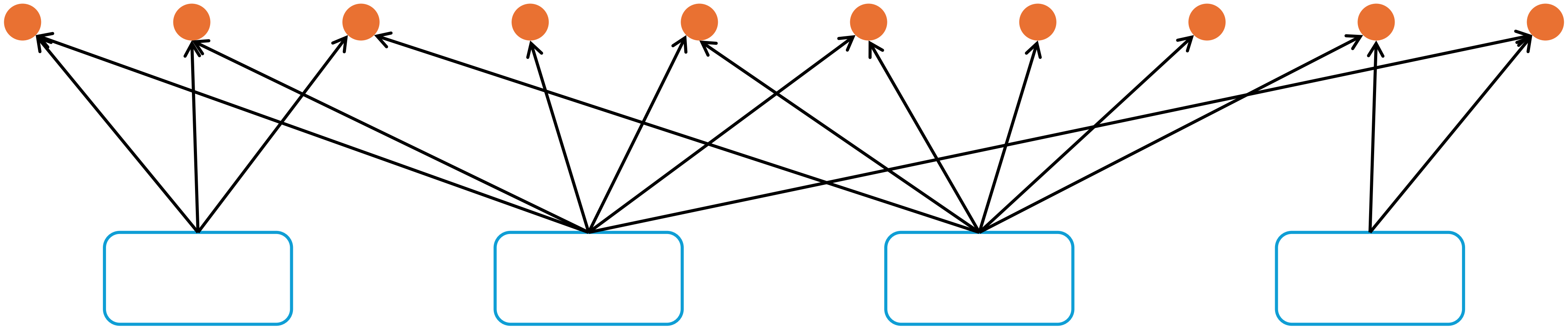
Private types: Model and results extend if there is a total ordering on receiver belief/type pairs that is monotone in action.

partition queries.

Model: Given a query consisting of a partition Q of beliefs, the oracle returns the subset $q \in Q$ containing the receiver's belief.

Theorem. Finding the optimal query policy is NP-complete.

Proof. A reduction from set-cover.

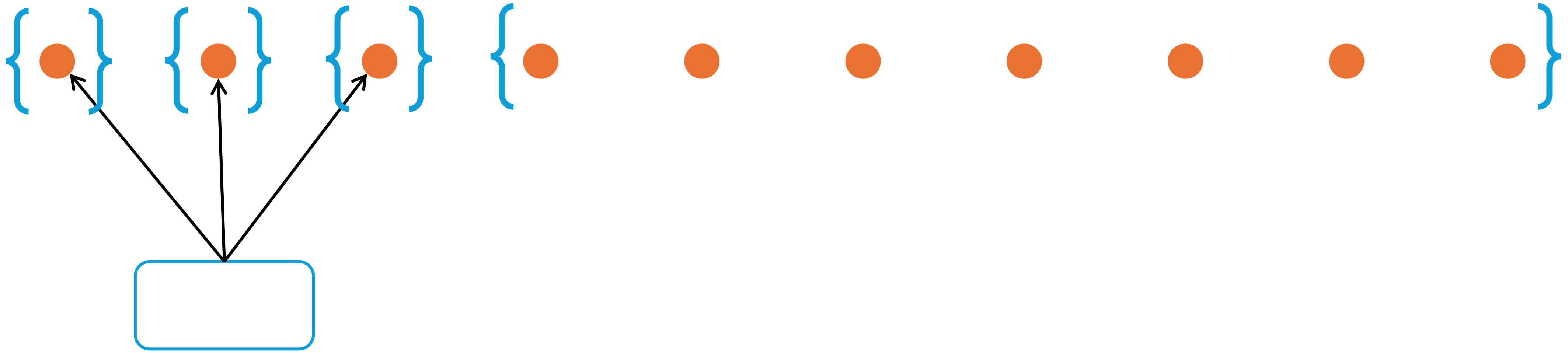


partition queries.

Model: Given a query consisting of a partition Q of beliefs, the oracle returns the subset $q \in Q$ containing the receiver's belief.

Theorem. Finding the optimal query policy is NP-complete.

Proof. A reduction from set-cover.

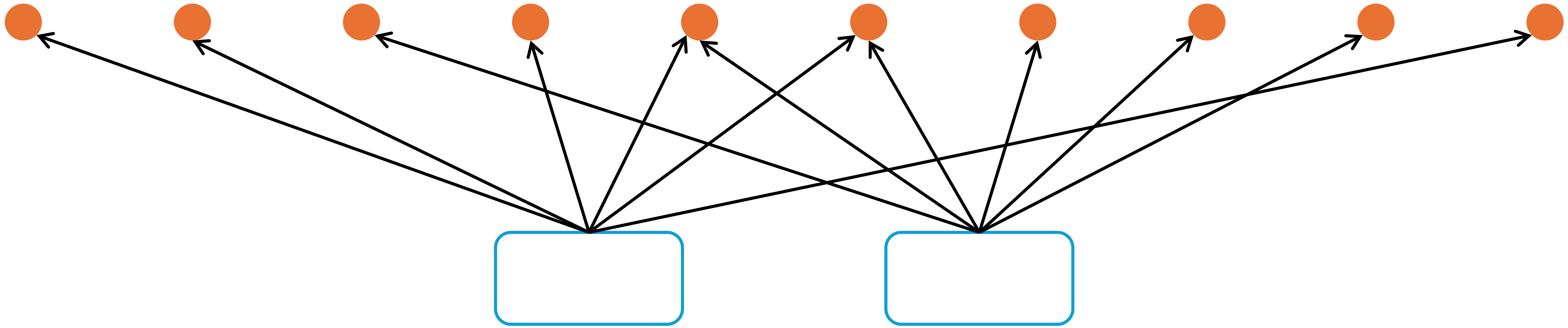


partition queries.

Model: Given a query consisting of a partition Q of beliefs, the oracle returns the subset $q \in Q$ containing the receiver's belief.

Theorem. Finding the optimal query policy is NP-complete.

Proof. A reduction from set-cover.



persuasion through simulation.

How can human agents leverage generative AI* to shape strategy?

Model: a binary action persuasion game where
- Receivers have additional signals of product quality
- AI simulates receiver choice for any sender messaging policy

Results:
- AI equivalent to a separation oracle on receiver beliefs
- An efficient algorithm for optimal query policy
- Extensions including error tolerance, private types, costly queries

* What if AI has its own incentives that are misaligned with those of its human user?