Dynamically Aggregating Diverse Information

Annie Liang[1] Xiaosheng Mu[2] Vasilis Syrgkanis[3]

[1]Northwestern

[2]Princeton [3]Microsoft Research

Simons Institute

Introduction

how should an agent acquire information **over time** given

- limited resources, and
- access to multiple kinds of information?

Introduction

how should an agent acquire information **over time** given

- limited resources, and
- access to multiple kinds of information?

examples:

- mayor wants to learn the COVID incidence rate in city, allocates limited number of tests across neighborhoods

Introduction

how should an agent acquire information **over time** given

- limited resources, and
- access to multiple kinds of information?

examples:

- mayor wants to learn the COVID incidence rate in city, allocates limited number of tests across neighborhoods
- news reader wants to learn the unknown cost of a proposed policy, allocates time across different (biased) news sources

This Talk

- model of the dynamic information acquisition problem
- main result: optimal information acquisition strategy can be exactly characterized and has an easily describable structure
- tractability of the model lends itself to application
- characterization can be used to derive new results in three settings motivated by particular economic questions

Model

Underlying Unknowns

unknown attributes $(\theta_1, \ldots, \theta_K) \sim \mathcal{N}(\mu, \Sigma)$

- e.g. each "attribute" is the COVID incidence rate in a specific neighborhood
- attributes may be correlated
- learn about θ_i by observing diffusion process X_i^t (more soon)

Underlying Unknowns

unknown attributes $(\theta_1, \ldots, \theta_K) \sim \mathcal{N}(\mu, \Sigma)$

- e.g. each "attribute" is the COVID incidence rate in a specific neighborhood
- attributes may be correlated
- learn about θ_i by observing diffusion process X_i^t (more soon)

payoff-relevant state: $\omega = \sum_{k=1}^{K} \alpha_k \theta_k$

- e.g. aggregate COVID incidence rate in city
- assume weights α_k are known

Attention Allocation

at each $t \in \mathbb{R}_+$, allocate budget of resources across attributes:

- choose $(\beta_1^t, \ldots, \beta_K^t)$ subject to $\beta_1^t + \cdots + \beta_K^t = 1$
- diffusion processes evolve as

$$dX_i^t = \beta_i^t \cdot \theta_i \cdot dt + \sqrt{\beta_i^t} \cdot dB_i^t$$

  where B_i are independent standard Brownian motions.

- more resources $\Rightarrow$ more precise information

Attention Allocation

at each $t \in \mathbb{R}_+$, allocate budget of resources across attributes:

- choose $(\beta_1^t, \ldots, \beta_K^t)$ subject to $\beta_1^t + \cdots + \beta_K^t = 1$
- diffusion processes evolve as

  $$dX_i^t = \beta_i^t \cdot \theta_i \cdot dt + \sqrt{\beta_i^t} \cdot dB_i^t$$

  where B_i are independent standard Brownian motions.
- more resources $\Rightarrow$ more precise information

discrete-time analogue: at each time $t \in \mathbb{Z}_+$, choose attention vector $(\beta_1(t), \ldots, \beta_K(t))$ summing to 1, and observe

$$\theta_i + \mathcal{N}\left(0, \frac{1}{\beta_i(t)}\right) \quad \text{for each } i = 1, \ldots, K$$

Decision Problem

- observe complete path of each process
- at each time t the history is $\left\{X_i^{\leq t}\right\}_{i=1}^K$
  - **information acquisition strategy** S: map from histories into an attention vector
  - **stopping rule** τ: map from history into decision of whether to stop sampling
- at endogenously chosen end time τ, take action $a \in A$ and receive $u(a, \omega, \tau)$

Related Literature

- not a multi-armed bandit problem (Gittins, 1979)

Related Literature

- not a multi-armed bandit problem (Gittins, 1979)
  - but related to "best-arm identification" when $K = 2$ (Bubeck et al. ('09); Russo ('16))
  - Frazier et al. ('08) show that the myopic "knowledge gradient policy" is optimal for two arms with independent payoffs

$\longrightarrow$ we consider many correlated unknowns that are aggregated to a one-dimensional payoff-relevant state

Related Literature

- not a multi-armed bandit problem (Gittins, 1979)
  - but related to "best-arm identification" when $K = 2$ (Bubeck et al. ('09); Russo ('16))
  - Frazier et al. ('08) show that the myopic "knowledge gradient policy" is optimal for two arms with independent payoffs

  $\longrightarrow$ we consider many correlated unknowns that are aggregated to a one-dimensional payoff-relevant state

- dynamic learning from fixed set of signals:
  - Fudenberg et al. ('18), Che and Mierendorff ('19); Mayskaya ('19); Gossner et al. ('20); Azevedo et al. ('20)
    $\longrightarrow$ we allow many signals with flexible correlation

Related Literature

- not a multi-armed bandit problem (Gittins, 1979)
  - but related to "best-arm identification" when $K = 2$ (Bubeck et al. ('09); Russo ('16))
  - Frazier et al. ('08) show that the myopic "knowledge gradient policy" is optimal for two arms with independent payoffs

  $\longrightarrow$ we consider many correlated unknowns that are aggregated to a one-dimensional payoff-relevant state

- dynamic learning from fixed set of signals:
  - Fudenberg et al. ('18), Che and Mierendorff ('19); Mayskaya ('19); Gossner et al. ('20); Azevedo et al. ('20)
    $\longrightarrow$ we allow many signals with flexible correlation
  - Callender ('11); Garfargnini and Strulovici ('16); Bardhi ('20)
    $\longrightarrow$ we have a finite number of attributes and noisy observations

Main Results:

Characterization of the Optimal Information Acquisition Strategy

Thm 1: result for $K = 2$

Thm 2: result for $K > 2$

Case of $K = 2$

- two attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

- payoff-relevant state is $\omega = \alpha_1\theta_1 + \alpha_2\theta_2$, where each $\alpha_i > 0$

Case of $K = 2$

- two attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

- payoff-relevant state is $\omega = \alpha_1\theta_1 + \alpha_2\theta_2$, where each $\alpha_i > 0$

- define $cov_i := \mathrm{Cov}(\omega, \theta_i) = \alpha_i\Sigma_{ii} + \alpha_j\Sigma_{ji}$ for each $i = 1, 2$

Assumption ("Attributes are Not Too Negatively Correlated")

$$cov_1 + cov_2 = \alpha_1\Sigma_{11} + \alpha_2\Sigma_{12} + \alpha_1\Sigma_{21} + \alpha_2\Sigma_{22} \geq 0$$

Case of $K = 2$

- two attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$$

- payoff-relevant state is $\omega = \alpha_1\theta_1 + \alpha_2\theta_2$, where each $\alpha_i > 0$
- define $cov_i := \mathrm{Cov}(\omega, \theta_i) = \alpha_i\Sigma_{ii} + \alpha_j\Sigma_{ji}$ for each $i = 1, 2$

Assumption ("Attributes are Not Too Negatively Correlated")

$$cov_1 + cov_2 = \alpha_1\Sigma_{11} + \alpha_2\Sigma_{12} + \alpha_1\Sigma_{21} + \alpha_2\Sigma_{22} \geq 0$$

sufficient conditions:

$$\alpha_1 = \alpha_2$$

Case of $K = 2$

- two attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

- payoff-relevant state is $\omega = \alpha_1\theta_1 + \alpha_2\theta_2$, where each $\alpha_i > 0$
- define $cov_i := \mathrm{Cov}(\omega, \theta_i) = \alpha_i\Sigma_{ii} + \alpha_j\Sigma_{ji}$ for each $i = 1, 2$

Assumption ("Attributes are Not Too Negatively Correlated")

$$cov_1 + cov_2 = \alpha_1\Sigma_{11} + \alpha_2\Sigma_{12} + \alpha_1\Sigma_{21} + \alpha_2\Sigma_{22} \geq 0$$

sufficient conditions:

$\alpha_1 = \alpha_2$ $\quad\quad$ $\Sigma_{12} = \Sigma_{21} \geq 0$

Case of $K = 2$

- two attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

- payoff-relevant state is $\omega = \alpha_1\theta_1 + \alpha_2\theta_2$, where each $\alpha_i > 0$
- define $cov_i := \mathrm{Cov}(\omega, \theta_i) = \alpha_i\Sigma_{ii} + \alpha_j\Sigma_{ji}$ for each $i = 1, 2$

Assumption ("Attributes are Not Too Negatively Correlated")

$$cov_1 + cov_2 = \alpha_1\Sigma_{11} + \alpha_2\Sigma_{12} + \alpha_1\Sigma_{21} + \alpha_2\Sigma_{22} \geq 0$$

sufficient conditions:

$$\alpha_1 = \alpha_2 \qquad \Sigma_{12} = \Sigma_{21} \geq 0 \qquad \Sigma_{11} = \Sigma_{22}$$

Optimal Attention Allocation Strategy

Theorem

Wlog let $cov_1 \geq cov_2$. Define

$$t_1 = \frac{cov_1 - cov_2}{\alpha_2 \det(\Sigma)}.$$

Optimal Attention Allocation Strategy

Theorem

Wlog let $cov_1 \geq cov_2$. Define

$$t_1 = \frac{cov_1 - cov_2}{\alpha_2 \det(\Sigma)}.$$

The optimal attention strategy has two stages:

Optimal Attention Allocation Strategy

Theorem

Wlog let $cov_1 \geq cov_2$. Define

$$t_1 = \frac{cov_1 - cov_2}{\alpha_2 \det(\Sigma)}.$$

The optimal attention strategy has two stages:

1. *At times $t \leq t_1$, DM allocates all attention to attribute 1.*

Optimal Attention Allocation Strategy

Theorem

Wlog let $cov_1 \geq cov_2$. Define

$$t_1 = \frac{cov_1 - cov_2}{\alpha_2 \det(\Sigma)}.$$

The optimal attention strategy has two stages:

1. *At times $t \leq t_1$, DM allocates all attention to attribute 1.*
2. *At times $t > t_1$, DM allocates attention in the constant fraction*

$$(\beta_1^t, \beta_2^t) = \left(\frac{\alpha_1}{\alpha_1 + \alpha_2}, \frac{\alpha_2}{\alpha_1 + \alpha_2} \right).$$

Example 1: Independent Attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

- payoff-relevant state is $\theta_1 + \theta_2$
- then optimally:
  - phase 1: put all attention on learning about θ_1
  - at time $t = 5/6$, posterior covariance matrix is $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
  - after, split attention equally

Example 2: Correlated Attributes

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 6 & 2 \\ 2 & 1 \end{pmatrix} \right)$$

- payoff-relevant state is $\theta_1 + \theta_2$
- then optimally:
  - phase 1: put all attention on learning about θ_1
  - at $t = 5/2$, posterior covariance is $\begin{pmatrix} 3/8 & 1/8 \\ 1/8 & 3/8 \end{pmatrix}$
  - after, split attention equally

$K > 2$ Attributes

Three different sufficient conditions (only need one):

- **Assumption 1:** (Perpetual Substitutes.) Σ^{-1} has negative off-diagonal entries.

- **Assumption 2:** (Perpetual Complements.) Σ has negative off-diagonal entries and $Cov(\theta_i, \omega) \geq 0$ for each attribute i.

- **Assumption 3:** (Diagonal Dominance.) Σ^{-1} is diagonally-dominant: $[\Sigma^{-1}]_{ii} \geq \sum_{j \neq i} |[\Sigma^{-1}]_{ij}| \; \forall \; i$.

$K > 2$ Attributes

Three different sufficient conditions (only need one):

- **Assumption 1:** (Perpetual Substitutes.) Σ^{-1} has negative off-diagonal entries.

  the partial correlation between any pair of attributes (controlling for all other attributes) is positive

- **Assumption 2:** (Perpetual Complements.) Σ has negative off-diagonal entries and $Cov(\theta_i, \omega) \geq 0$ for each attribute i.

- **Assumption 3:** (Diagonal Dominance.) Σ^{-1} is diagonally-dominant: $[\Sigma^{-1}]_{ii} \geq \sum_{j \neq i} |[\Sigma^{-1}]_{ij}| \ \forall \ i$.

$K > 2$ Attributes

Three different sufficient conditions (only need one):

- **Assumption 1:** (Perpetual Substitutes.) Σ^{-1} has negative off-diagonal entries.

  the partial correlation between any pair of attributes (controlling for all other attributes) is positive

- **Assumption 2:** (Perpetual Complements.) Σ has negative off-diagonal entries and $Cov(\theta_i, \omega) \geq 0$ for each attribute i.

  prior covariances are mildly negative

- **Assumption 3:** (Diagonal Dominance.) Σ^{-1} is diagonally-dominant: $[\Sigma^{-1}]_{ii} \geq \sum_{j \neq i} |[\Sigma^{-1}]_{ij}| \; \forall \; i$.

$K > 2$ Attributes

Three different sufficient conditions (only need one):

- **Assumption 1:** (Perpetual Substitutes.) Σ^{-1} has negative off-diagonal entries.

  the partial correlation between any pair of attributes (controlling for all other attributes) is positive

- **Assumption 2:** (Perpetual Complements.) Σ has negative off-diagonal entries and $Cov(\theta_i, \omega) \geq 0$ for each attribute i.

  prior covariances are mildly negative

- **Assumption 3:** (Diagonal Dominance.) Σ^{-1} is diagonally-dominant: $[\Sigma^{-1}]_{ii} \geq \sum_{j \neq i} |[\Sigma^{-1}]_{ij}| \ \forall \ i$.

  covariance matrix is not too far from identity

Optimal Information Acquisition Strategy

Theorem

Under any of the preceding assumptions, there exist times

$$0 = t_0 < t_1 < \cdots < t_m = +\infty$$

and nested sets

$$\emptyset \subsetneq B_1 \subsetneq \cdots \subsetneq B_m = \{1, \ldots, K\},$$

such that an optimal information acquisition strategy is described by a deterministic path of attention allocations.

Optimal Information Acquisition Strategy

Theorem

Under any of the preceding assumptions, there exist times

$$0 = t_0 < t_1 < \cdots < t_m = +\infty$$

and nested sets

$$\emptyset \subsetneq B_1 \subsetneq \cdots \subsetneq B_m = \{1, \ldots, K\},$$

such that an optimal information acquisition strategy is described by a deterministic path of attention allocations.

At each stage $[t_{k-1}, t_k)$:

- *the optimal attention level is constant*
- *and supported on the sources in B_k.*

Optimal Information Acquisition Strategy

Theorem

Under any of the preceding assumptions, there exist times

$$0 = t_0 < t_1 < \cdots < t_m = +\infty$$

and nested sets

$$\emptyset \subsetneq B_1 \subsetneq \cdots \subsetneq B_m = \{1, \ldots, K\},$$

such that an optimal information acquisition strategy is described by a deterministic path of attention allocations.

At each stage $[t_{k-1}, t_k)$*:*

- *the optimal attention level is constant*
- *and supported on the sources in* B_k*.*

At the final stage, attention is proportional to the weight vector α*.*

Optimal Information Acquisition Strategy

Theorem

Under any of the preceding assumptions, there exist times

$$0 = t_0 < t_1 < \cdots < t_m = +\infty$$

and nested sets

$$\emptyset \subsetneq B_1 \subsetneq \cdots \subsetneq B_m = \{1, \ldots, K\},$$

such that an optimal information acquisition strategy is described by a deterministic path of attention allocations.

At each stage $[t_{k-1}, t_k)$:

- *the optimal attention level is constant*
- *and supported on the sources in B_k.*

At the final stage, attention is proportional to the weight vector α.

- full path can be computed from α and Σ (see paper)

Properties of the Solution

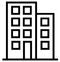
The optimal attention allocation strategy is:

- history-independent (can map out full path from $t = 0$)
- independent of the stopping rule
  - don't have to solve for stopping rule and information acquisition strategy jointly
- robust across decision problems

Explanation of Results

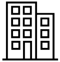
Static Problem

one-time budget of t **total tests**



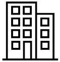
Testing Center 1

θ_1



Testing Center 2

θ_2



Testing Center 3

θ_3

posterior variance of ω can be written as a function $V(q_1, q_2, q_3)$

static problem: choose $q_1, q_2, q_3 \in \mathbb{R}_+$ to minimize $V(q_1, q_2, q_3)$
subject to $q_1 + q_2 + q_3 \leq t$

Static Problem

one-time budget of t **total tests**

Testing Center 1

θ_1

optimally allocate $q_1^*(t)$ tests

Testing Center 2

θ_2

optimally allocate $q_2^*(t)$ test

Testing Center 3

θ_3

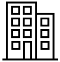
optimally allocate $q_3^*(t)$ tests

posterior variance of ω can be written as a function $V(q_1, q_2, q_3)$

static problem: choose $q_1, q_2, q_3 \in \mathbb{R}_+$ to minimize $V(q_1, q_2, q_3)$
subject to $q_1 + q_2 + q_3 \leq t$

Exogenous End Time $T = 100$

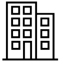
100 total tests



Testing Center 1

θ_1

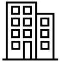
100 tests



Testing Center 2

θ_2

0 tests



Testing Center 3

θ_3

0 tests

Exogenous End Time $T = 101$

101 total tests

Testing Center 1
θ_1
1 test

Testing Center 2
θ_2
50 tests

Testing Center 3
θ_3
50 tests

Exogenous End Time $T = 101$

101 total tests

Testing Center 1
θ_1

1 test

Testing Center 2
θ_2

50 tests

Testing Center 3
θ_3

50 tests

DM faces intertemporal tradeoffs: must choose between better information for a decision at time $t = 100$ versus $t = 101$

Key Idea: Uniformly Optimal Strategies

- Iff $\boldsymbol{q}^*(t)$ is increasing in in each of its coordinates, possible to achieve $\boldsymbol{q}^*(t)$ at every t along a single sampling strategy

Key Idea: Uniformly Optimal Strategies

- Iff $\boldsymbol{q}^*(t)$ is increasing in in each of its coordinates, possible to achieve $\boldsymbol{q}^*(t)$ at every t along a single sampling strategy

- Call such a strategy **uniformly optimal**.
    - minimizes posterior variance at every moment
    - **lemma:** best for all decision problems

Key Idea: Uniformly Optimal Strategies

- Iff $\boldsymbol{q}^*(t)$ is increasing in in each of its coordinates, possible to achieve $\boldsymbol{q}^*(t)$ at every t along a single sampling strategy

- Call such a strategy **uniformly optimal**.
  - minimizes posterior variance at every moment
  - **lemma:** best for all decision problems

- Our different sufficient conditions on the prior guarantee that $\boldsymbol{q}^*(t)$ is increasing in t

When Does a Uniformly Optimal Strategy Exist?

- When is $\boldsymbol{q}^*(t)$ increasing in t?

When Does a Uniformly Optimal Strategy Exist?

- When is $\boldsymbol{q}^*(t)$ increasing in t?

- Analogy with a classic consumer demand theory problem:
  - Utility function $U(q_1, \ldots, q_K)$ over consumption of q_k units of each of K goods
  - Let $D(\boldsymbol{p}, w)$ denote consumer's demand subject to budget constraint $\boldsymbol{p} \cdot \boldsymbol{q} \leq w$.
  - Demand is **normal** if each coordinate of $D(\boldsymbol{p}, w)$ increases with income w.

When Does a Uniformly Optimal Strategy Exist?

- When is $\boldsymbol{q}^*(t)$ increasing in t?

- Analogy with a classic consumer demand theory problem:
  - Utility function $U(q_1, \ldots, q_K)$ over consumption of q_k units of each of K goods
  - Let $D(\boldsymbol{p}, w)$ denote consumer's demand subject to budget constraint $\boldsymbol{p} \cdot \boldsymbol{q} \leq w$.
  - Demand is **normal** if each coordinate of $D(\boldsymbol{p}, w)$ increases with income w.

- Let $U = -V$, $\boldsymbol{p} = (1, 1, \ldots, 1)'$, and $w = t$. Then normality of demand is equivalent to monotonicity of $\boldsymbol{q}^*(t)$.

When Does a Uniformly Optimal Strategy Exist?

- When is $\boldsymbol{q}^*(t)$ increasing in t?
- Analogy with a classic consumer demand theory problem:
  - Utility function $U(q_1, \ldots, q_K)$ over consumption of q_k units of each of K goods
  - Let $D(\boldsymbol{p}, w)$ denote consumer's demand subject to budget constraint $\boldsymbol{p} \cdot \boldsymbol{q} \leq w$.
  - Demand is **normal** if each coordinate of $D(\boldsymbol{p}, w)$ increases with income w.
- Let $U = -V$, $\boldsymbol{p} = (1, 1, \ldots, 1)'$, and $w = t$. Then normality of demand is equivalent to monotonicity of $\boldsymbol{q}^*(t)$.
- Our condition "Perpetual Complementarity" is directly related to a sufficient condition for normality of demand.

When Does a Uniformly Optimal Strategy Exist?

- When is $\boldsymbol{q}^*(t)$ increasing in t?
- Analogy with a classic consumer demand theory problem:
  - Utility function $U(q_1, \ldots, q_K)$ over consumption of q_k units of each of K goods
  - Let $D(\boldsymbol{p}, w)$ denote consumer's demand subject to budget constraint $\boldsymbol{p} \cdot \boldsymbol{q} \leq w$.
  - Demand is **normal** if each coordinate of $D(\boldsymbol{p}, w)$ increases with income w.
- Let $U = -V$, $\boldsymbol{p} = (1, 1, \ldots, 1)'$, and $w = t$. Then normality of demand is equivalent to monotonicity of $\boldsymbol{q}^*(t)$.
- Our condition "Perpetual Complementarity" is directly related to a sufficient condition for normality of demand.
- We exploit properties of $U = -V$ to derive the others.

Structure of Uniformly Optimal Strategies

- The attention allocations β^t under the uniformly optimal strategy are simply the time derivatives of $\boldsymbol{q}^*(t)$.
  - i.e. "greedy" optimization

Structure of Uniformly Optimal Strategies

- The attention allocations β^t under the uniformly optimal strategy are simply the time derivatives of $\boldsymbol{q}^*(t)$.
    - i.e. "greedy" optimization

- At each stage, agent optimally divides attention among the set of attributes with highest marginal value for learning about ω.

Structure of Uniformly Optimal Strategies

- The attention allocations β^t under the uniformly optimal strategy are simply the time derivatives of $\boldsymbol{q}^*(t)$.
  - i.e. "greedy" optimization

- At each stage, agent optimally divides attention among the set of attributes with highest marginal value for learning about ω.

- At each stage, the mixture maintains equivalence of marginal values of those attributes, but reduces it.

Structure of Uniformly Optimal Strategies

- The attention allocations β^t under the uniformly optimal strategy are simply the time derivatives of $\boldsymbol{q}^*(t)$.
    - i.e. "greedy" optimization
- At each stage, agent optimally divides attention among the set of attributes with highest marginal value for learning about ω.
- At each stage, the mixture maintains equivalence of marginal values of those attributes, but reduces it.
- Eventually, some other attribute has the same marginal value and the agent expands his observation set to include it. Etc.

Application of Characterization

- Can apply characterizations to derive new results in settings motivated by particular economic questions.

- We illustrate this with three applications, where we use our main results to:
  - tractably introduce correlation in settings that have been previously studied under strong assumptions of independence.
  - derive results about other economic behaviors.

Summary of Application 1: Binary Choice

- DM learns about unknown payoffs $(v_1, v_2) \sim \mathcal{N}(\mu, \Sigma)$ of two goods before making a choice.
- Set $\theta_1 = v_1$, $\theta_2 = -v_2$, $\omega = \theta_1 + \theta_2$ and observe that one of the sufficient conditions for $K = 2$ is met $(\alpha_1 = \alpha_2)$.
- So our main result yields the optimal information acquisition strategy.
- Use this to generalize a result from Fudenberg et al. ('18) regarding the relationship between choice speed and accuracy.

Summary of Application 2: Attention Manipulation

- Gossner et al. ('21) study the dynamic implications of attention manipulation in a model with goods with independent payoffs.
- Diverting attention towards a specific good leads to
  - persistently higher cumulative attention devoted to that good
  - persistently lower cumulative attention to every other good
- We derive a complementary result in our setting, focusing on the role of correlation:
  - Gossner et al. ('21)'s qualitative conclusion can in general fail with correlation
  - But extends under the "Perpetual Substitutes" condition identified earlier

Summary of Application 3: Biased News Sources

- Stylized game between a liberal and a conservative news source
  - Report on a common unknown (e.g., the fiscal cost of a policy proposal), but reporting is biased in opposite directions.
  - Sources choose the size of their bias and the precision of their reporting, and compete over readers' attention.
- Apply our result to characterize equilibrium news provision in this model.
- Find that higher intrinsic incentives for bias not only lead to greater polarization in equilibrium, but also lead to less precise reporting.

Conclusion

- Information acquisition is a classic problem within economics, but relatively few dynamic models are simultaneously rich and tractable.
- We present a class of dynamic information acquisition problems whose solution can be explicitly characterized in closed form.
- Key restrictions:
  - Gaussian uncertainty
  - a one-dimensional payoff-relevant state
  - correlation across the unknowns that satisfies certain assumptions (e.g., if correlation is not too strong)
- Can accommodate generality in other aspects of the problem (e.g., the decision problem and the agent's time preferences)
- The tractability of the solution and the flexibility of the environment open the door to interesting applications.

Thank You!