

Dynamically Aggregating Diverse Information

Annie Liang¹ Xiaosheng Mu² Vasilis Syrgkanis³

¹Northwestern

²Princeton ³Microsoft Research

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 $(\theta_1, \dots, \theta_K) \sim \mathcal{N}(\mu, \Sigma)$

- 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 $(\theta_1, \dots, \theta_K) \sim \mathcal{N}(\mu, \Sigma)$

- 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 $(\beta_1^t, \dots, \beta_K^t)$ subject to $\beta_1^t + \dots + \beta_K^t = 1$
- 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 $(\beta_1^t, \dots, \beta_K^t)$ subject to $\beta_1^t + \dots + \beta_K^t = 1$
- 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 $(\beta_1(t), \dots, \beta_K(t))$ summing to 1, and observe

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

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)
 - recent work on dynamic learning from fixed set of signals:
 - Fudenberg, Strack, and Strzalecki ('18), Che and Mierendorff ('19); Mayskaya ('19); Bardhi ('20); Gossner, Steiner, and Stewart ('20); Azevedo et al ('20)
- we allow many signals with flexible correlation

Related Literature

- not a multi-armed bandit problem (Gittins, 1979)
- recent work on dynamic learning from fixed set of signals:
 - Fudenberg, Strack, and Strzalecki ('18), Che and Mierendorff ('19); Mayskaya ('19); Bardhi ('20); Gossner, Steiner, and Stewart ('20); Azevedo et al ('20)

→ we allow many signals with flexible correlation
- rational inattention and flexible information acquisition:
 - Steiner, Stewart, and Matejka ('17); Hébert and Woodford ('19); Morris and Strack ('19); Zhong ('19)

→ our signals and information cost are prior-independent

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 $\alpha_j > 0$

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 $cov_i := \text{Cov}(\omega, \theta_i) = \alpha_i\Sigma_{ii} + \alpha_j\Sigma_{ji}$ 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 $cov_i := \text{Cov}(\omega, \theta_i) = \alpha_i\Sigma_{ii} + \alpha_j\Sigma_{ji}$ 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 $cov_i := \text{Cov}(\omega, \theta_i) = \alpha_i\Sigma_{ii} + \alpha_j\Sigma_{ji}$ 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:

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

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 $cov_i := \text{Cov}(\omega, \theta_i) = \alpha_i\Sigma_{ii} + \alpha_j\Sigma_{ji}$ 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:

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

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 $\text{cov}_1 \geq \text{cov}_2$. Define

$$t_1 = \frac{\text{cov}_1 - \text{cov}_2}{\alpha_2 \det(\Sigma)}.$$

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 $\text{Cov}(\theta_i, \omega) \geq 0$ 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 $\text{Cov}(\theta_i, \omega) \geq 0$ 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

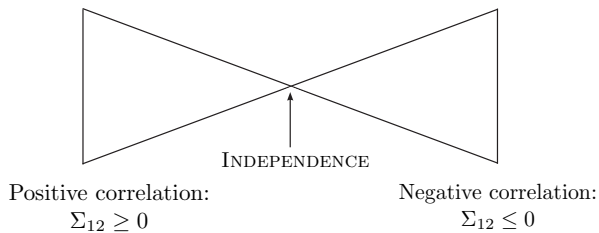
- **Assumption 2:** (Perpetual Complements.) Σ has negative off-diagonal entries and $\text{Cov}(\theta_i, \omega) \geq 0$ for each attribute i .

prior covariances are mildly negative

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

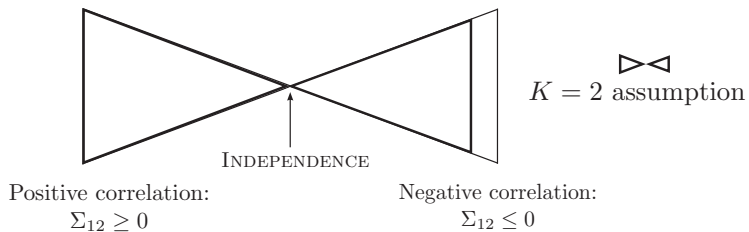
covariance matrix is not too far from identity

Relationship Between Conditions



$$K = 2$$

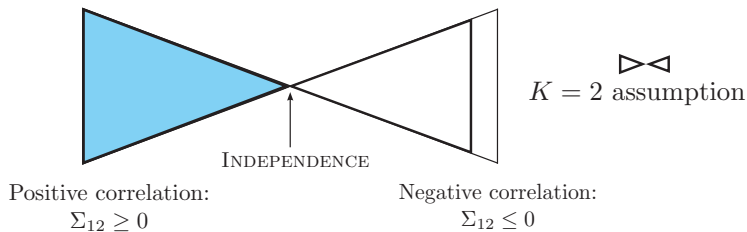
Relationship Between Conditions



$$K = 2$$

Relationship Between Conditions

■ Perpetual Substitutes: $[\Sigma^{-1}]_{ij} \leq 0$ for all $i \neq j$

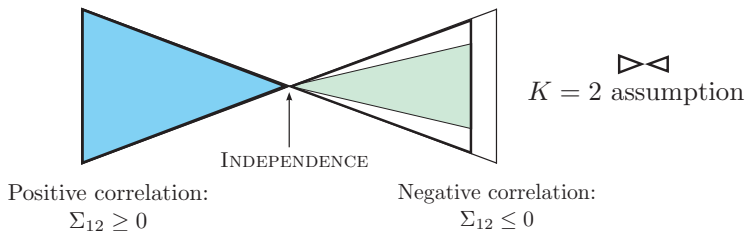


$$K = 2$$

Relationship Between Conditions

■ Perpetual Substitutes: $[\Sigma^{-1}]_{ij} \leq 0$ for all $i \neq j$

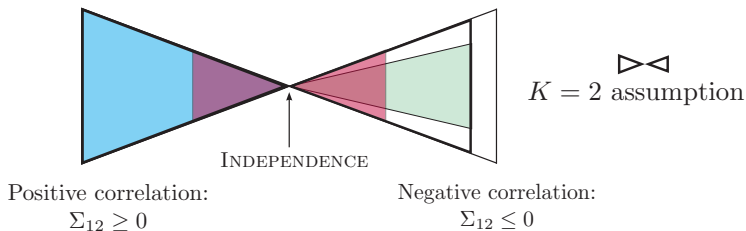
■ Perpetual Complements: $\Sigma_{ij} \leq 0$ for all $i \neq j$ and $\Sigma \cdot \alpha \in \mathbb{R}_+^K$



$$K = 2$$

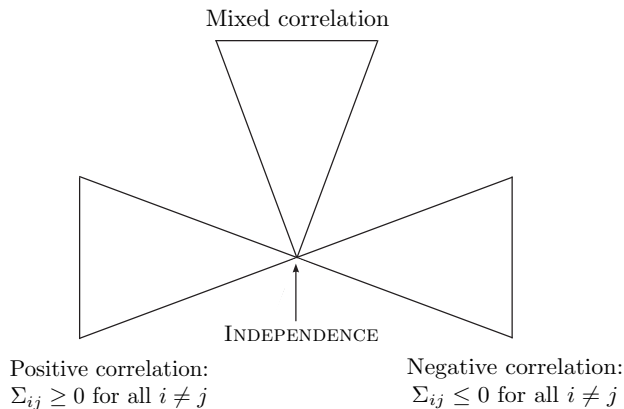
Relationship Between Conditions

- Perpetual Substitutes: $[\Sigma^{-1}]_{ij} \leq 0$ for all $i \neq j$
- Perpetual Complements: $\Sigma_{ij} \leq 0$ for all $i \neq j$ and $\Sigma \cdot \alpha \in \mathbb{R}_+^K$
- Diagonal Dominance: Σ^{-1} is diagonally dominant



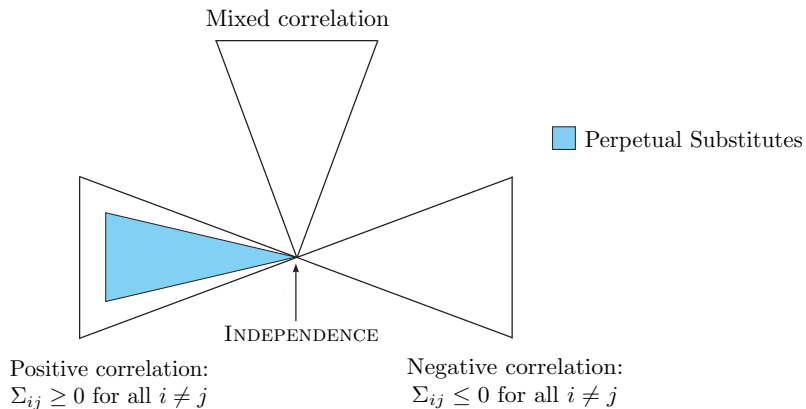
$$K = 2$$

Relationship Between Conditions



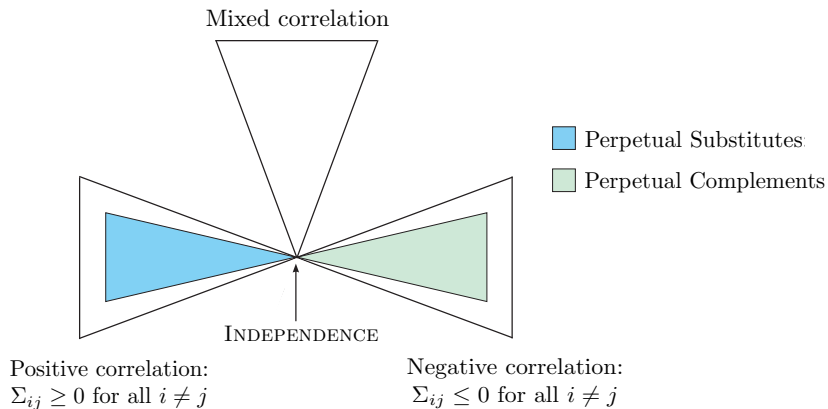
$$K > 2$$

Relationship Between Conditions



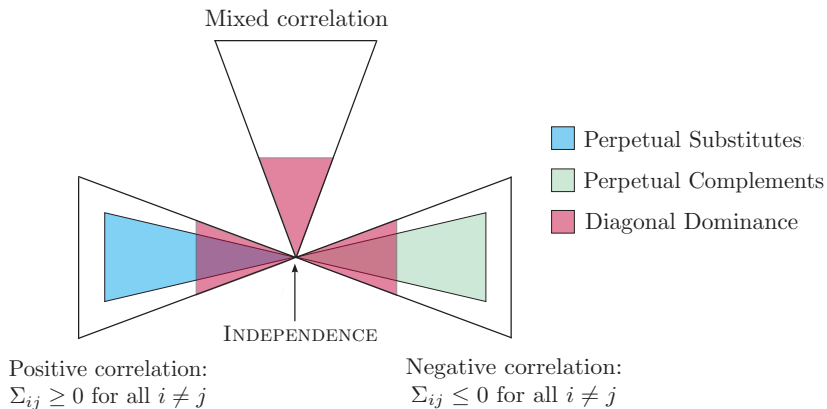
$$K > 2$$

Relationship Between Conditions



$$K > 2$$

Relationship Between Conditions



$$K > 2$$

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

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

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

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

and nested sets

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

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

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

and nested sets

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

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

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

and nested sets

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

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 $q_2^*(t)$ tests



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

Key Idea: Uniformly Optimal Strategies

- Iff $\mathbf{q}^*(t)$ is increasing in each of its coordinates, possible to achieve $\mathbf{q}^*(t)$ 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 $\mathbf{q}^*(t)$ is increasing in each of its coordinates, possible to achieve $\mathbf{q}^*(t)$ 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 $\mathbf{q}^*(t)$ is increasing in t

When Does a Uniformly Optimal Strategy Exist?

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

When Does a Uniformly Optimal Strategy Exist?

- When is $\mathbf{q}^*(t)$ increasing in t ?
- Analogy with a classic consumer demand theory problem:
 - Utility function $U(q_1, \dots, q_K)$ over consumption of q_k units of each of K goods
 - Let $D(\mathbf{p}, w)$ denote consumer's demand subject to budget constraint $\mathbf{p} \cdot \mathbf{q} \leq w$.
 - Demand is **normal** if each coordinate of $D(\mathbf{p}, w)$ increases with income w .

When Does a Uniformly Optimal Strategy Exist?

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

When Does a Uniformly Optimal Strategy Exist?

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

When Does a Uniformly Optimal Strategy Exist?

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

Application 1: Binary Choice

Uncertain Drift Diffusion Model

Fudenberg, Strack, and Strzalecki (2018) recently proposed the **uncertain drift diffusion model**:

- Agent has choice between two goods with unknown payoffs

$$(v_1, v_2)' \sim \mathcal{N} \left((\mu_1, \mu_2)', \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \right)$$

- Agent continuously divides a unit of attention across two Brownian processes whose drifts are the unknown payoffs.
- The agent chooses a stopping time τ to maximize

$$\mathbb{E}[\mathbb{E}[\max\{v_1, v_2\} \mid \mathcal{F}_\tau] - c\tau],$$

where $c\tau$ is a linear waiting cost.

Relationship to Our Framework

- The payoff difference $v_1 - v_2$ is a sufficient statistic for the agent's decision.

Relationship to Our Framework

- The payoff difference $v_1 - v_2$ is a sufficient statistic for the agent's decision.
- Define $\theta_1 = v_1$ and $\theta_2 = -v_2$. Then the FSS model is nested in our framework with $\omega = \theta_1 + \theta_2$.

Relationship to Our Framework

- The payoff difference $v_1 - v_2$ is a sufficient statistic for the agent's decision.
- Define $\theta_1 = v_1$ and $\theta_2 = -v_2$. Then the FSS model is nested in our framework with $\omega = \theta_1 + \theta_2$.
- Different from FSS, suppose

$$(v_1, v_2)' \sim \mathcal{N} \left((\mu_1, \mu_2)', \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \right)$$

Relationship to Our Framework

- The payoff difference $v_1 - v_2$ is a sufficient statistic for the agent's decision.
- Define $\theta_1 = v_1$ and $\theta_2 = -v_2$. Then the FSS model is nested in our framework with $\omega = \theta_1 + \theta_2$.
- Different from FSS, suppose

$$(v_1, v_2)' \sim \mathcal{N} \left((\mu_1, \mu_2)', \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

Relationship to Our Framework

- The payoff difference $v_1 - v_2$ is a sufficient statistic for the agent's decision.
- Define $\theta_1 = v_1$ and $\theta_2 = -v_2$. Then the FSS model is nested in our framework with $\omega = \theta_1 + \theta_2$.

- Different from FSS, suppose

$$(v_1, v_2)' \sim \mathcal{N} \left((\mu_1, \mu_2)', \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

- Correlation and asymmetry both typical properties of choice environments:
 - Value of two stocks correlated by global economic shocks
 - Uncertainty about value of PC vs Mac depends on prior experience with either computer

Endogenous Allocation of Attention

Theorem 5, FSS (Optimal Endogenous Allocation of Attention)

Suppose $\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$. The agent optimally divides attention equally at every moment of time.

Endogenous Allocation of Attention

Theorem 5, FSS (Optimal Endogenous Allocation of Attention)

Suppose $\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$. The agent optimally divides attention equally at every moment of time.

Our Generalization

Suppose $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ with $\Sigma_{11} \geq \Sigma_{22}$.

- **Stage 1:** Prior to time $t_1^* = \frac{\Sigma_{11} - \Sigma_{22}}{\det(\Sigma)}$, the agent optimally allocates all attention to θ_1 .
- **Stage 2:** After t_1^* , the agent optimally allocates attention equally.

Endogenous Allocation of Attention

Theorem 5, FSS (Optimal Endogenous Allocation of Attention)

Suppose $\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$. The agent optimally divides attention equally at every moment of time.

Our Generalization

Suppose $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ with $\Sigma_{11} \geq \Sigma_{22}$.

- **Stage 1:** Prior to time $t_1^* = \frac{\Sigma_{11} - \Sigma_{22}}{\det(\Sigma)}$, the agent optimally allocates all attention to θ_1 .
- **Stage 2:** After t_1^* , the agent optimally allocates attention equally.
- Length of Stage 1, t_1^* , is increasing in asymmetry between initial uncertainty and correlation between the payoffs.

Earlier Decisions are More Accurate

Let $p(t)$ be the probability of choosing the higher-value good conditional on stopping at time t .

Earlier Decisions are More Accurate

Let $p(t)$ be the probability of choosing the higher-value good conditional on stopping at time t . Two opposing forces:

- More information at later times.
- More likely to stop earlier when the decision is easy.

Earlier Decisions are More Accurate

Let $p(t)$ be the probability of choosing the higher-value good conditional on stopping at time t . Two opposing forces:

- More information at later times.
- More likely to stop earlier when the decision is easy.

Proposition 3, FSS (Speed and Accuracy)

Suppose $\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$. $p(t)$ is (weakly) decreasing over time.

Earlier Decisions are More Accurate

Let $p(t)$ be the probability of choosing the higher-value good conditional on stopping at time t . Two opposing forces:

- More information at later times.
- More likely to stop earlier when the decision is easy.

Proposition 3, FSS (Speed and Accuracy)

Suppose $\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$. $p(t)$ is (weakly) decreasing over time.

Our Generalization

For any Σ , $p(t)$ is (weakly) decreasing over time.

Applying our Characterization

How we use the characterization of attention allocation in generalizing the result that earlier decisions are more accurate:

Recall that the characterization is:

- **Stage 1:** Prior to time $t_1^* = \frac{\Sigma_{11} - \Sigma_{22}}{\det(\Sigma)}$, the agent optimally allocates all attention to θ_1 .
- **Stage 2:** After t_1^* , the agent optimally allocates attention equally.

Applying our Characterization

How we use the characterization of attention allocation in generalizing the result that earlier decisions are more accurate:

Recall that the characterization is:

- **Stage 1:** Prior to time $t_1^* = \frac{\Sigma_{11} - \Sigma_{22}}{\det(\Sigma)}$, the agent optimally allocates all attention to θ_1 .
- **Stage 2:** After t_1^* , the agent optimally allocates attention equally.

→ optimal attention strategy doesn't depend on the stopping time

- can take information as given exogenously

Applying our Characterization

How we use the characterization of attention allocation in generalizing the result that earlier decisions are more accurate:

Recall that the characterization is:

- **Stage 1:** Prior to time $t_1^* = \frac{\Sigma_{11} - \Sigma_{22}}{\det(\Sigma)}$, the agent optimally allocates all attention to θ_1 .
- **Stage 2:** After t_1^* , the agent optimally allocates attention equally.

→ optimal attention strategy doesn't depend on the stopping time

- can take information as given exogenously

→ result pins down the evolution of Σ_t

- shows in particular that asymmetry in uncertainty is decreasing in time along the optimal path, which turns out to be key

Application 2:
Biased News Sources

The Setting

- Two sources $i = 1, 2$ (e.g. liberal and conservative) report on

$$\omega \sim \mathcal{N}(\mu_\omega, \sigma_\omega^2)$$

e.g., the cost of a proposed policy

- Partisan implications are not precisely known by the general public (although they are understood by the sources)
 - e.g. new limits on short selling in financial markets or trade deals with countries in Southeast Asia
- Define b to be the benefit to source 1's party when the reader believes that ω is large.
- From the perspective of the reader, b is a random variable with distribution $b \sim \mathcal{N}(\mu_b, \sigma_b^2)$

The Game

Sources bias their reporting in opposite directions:

- Unit of time on source 1

$$\hookrightarrow X_1 \sim \mathcal{N}(\omega + \phi_1 \mathbf{b}, \zeta_1^2)$$

- Unit of time on source 2

$$\hookrightarrow X_2 \sim \mathcal{N}(\omega - \phi_2 \mathbf{b}, \zeta_2^2)$$

where both the intensity of bias ($\phi_i > 0$) and noisiness of reporting ($\zeta_i > 0$) are choice variables.

Source Payoffs

A representative news reader faces a decision that depends on ω , and optimally allocates attention over time $\rightarrow (\beta_1^t, \beta_2^t)$.

Source Payoffs

A representative news reader faces a decision that depends on ω , and optimally allocates attention over time $\rightarrow (\beta_1^t, \beta_2^t)$.

Each source i 's payoff is the sum of discounted attention and a reward for bias:

$$U_i = \underbrace{\int_0^{\infty} r e^{-rt} \beta_i^t dt}_{\text{discounted average attention}} - \lambda(\phi_i - \kappa)^2.$$

discounted average attention

r is the discount rate

Source Payoffs

A representative news reader faces a decision that depends on ω , and optimally allocates attention over time $\rightarrow (\beta_1^t, \beta_2^t)$.

Each source i 's payoff is the sum of discounted attention and a reward for bias:

$$U_i = \int_0^{\infty} re^{-rt} \beta_i^t dt - \underbrace{\lambda(\phi_i - \kappa)^2}_{\text{reward for bias}}.$$

reward for bias

λ moderates strength of incentive
 κ is the bliss point for the bias intensity

Applying our Characterization

Fixing choices $\phi_1, \phi_2, \zeta_1, \zeta_2$ by the two sources, our characterization allows us to

Applying our Characterization

Fixing choices $\phi_1, \phi_2, \zeta_1, \zeta_2$ by the two sources, our characterization allows us to

→ pin down attention path $(\beta_1(t), \beta_2(t))$

Applying our Characterization

Fixing choices $\phi_1, \phi_2, \zeta_1, \zeta_2$ by the two sources, our characterization allows us to

→ pin down attention path $(\beta_1(t), \beta_2(t))$

→ derive payoffs $U_i(\phi_1, \phi_2; \zeta_1, \zeta_2)$

Applying our Characterization

Fixing choices $\phi_1, \phi_2, \zeta_1, \zeta_2$ by the two sources, our characterization allows us to

→ pin down attention path $(\beta_1(t), \beta_2(t))$

→ derive payoffs $U_i(\phi_1, \phi_2; \zeta_1, \zeta_2)$

→ solve for equilibrium

Equilibrium

Proposition

Suppose $\lambda\kappa^2 \geq 1.6$. The unique pure strategy equilibrium is $(\phi_1^, \zeta_1^*; \phi_2^*, \zeta_2^*)$ where*

$$\phi_1^* = \phi_2^* = \frac{1}{2} \left(\kappa + \sqrt{\kappa^2 - \frac{1}{2\lambda}} \right)$$

and

$$\zeta_1^* = \zeta_2^* = \frac{\sigma_b}{2\sqrt{r}} \cdot \left(\kappa + \sqrt{\kappa^2 - \frac{1}{2\lambda}} \right).$$

Given these equilibrium choices, the reader optimally devotes equal attention to the two sources at every moment.

Equilibrium News Informativeness

Corollary (Informativeness of News)

The equilibrium noise level ζ^ is*

- (a) increasing in the incentive for bias λ and the bliss point κ for the bias intensity;*

- (b) decreasing in the discount rate r .*

Equilibrium News Informativeness

$$X_1 \sim \mathcal{N}(\omega + \phi_1^* b, \zeta_1^{*2}) \quad X_2 \sim \mathcal{N}(\omega + \phi_2^* b, \zeta_2^{*2})$$

Corollary (Informativeness of News)

The equilibrium noise level ζ^ is*

- (a) *increasing in the incentive for bias λ and the bliss point κ for the bias intensity;*

→ incentives for bias not only increase polarization in equilibrium, but also decrease quality of reporting

- (b) *decreasing in the discount rate r .*

Equilibrium News Informativeness

$$X_1 \sim \mathcal{N}(\omega + \phi_1^* b, \zeta_1^{*2}) \quad X_2 \sim \mathcal{N}(\omega + \phi_2^* b, \zeta_2^{*2})$$

Corollary (Informativeness of News)

The equilibrium noise level ζ^ is*

- (a) *increasing in the incentive for bias λ and the bliss point κ for the bias intensity;*

→ incentives for bias not only increase polarization in equilibrium, but also decrease quality of reporting

- (b) *decreasing in the discount rate r .*

→ patient news sources provide lower quality news

Intuition

- Suppose $\phi_1 = \phi_2$. There are (up to) two stages of info acquisition:
 - Stage 1: source i with smaller noise ζ_i receives all attention
 - Stage 2: each source i receives fraction $\frac{\zeta_i}{\zeta_1 + \zeta_2}$

Intuition

- Suppose $\phi_1 = \phi_2$. There are (up to) two stages of info acquisition:
 - Stage 1: source i with smaller noise ζ_i receives all attention
 - Stage 2: each source i receives fraction $\frac{\zeta_i}{\zeta_1 + \zeta_2}$
- Firms face a tradeoff between optimizing for long-run viewership and competing to be chosen in the short run

Intuition

- Suppose $\phi_1 = \phi_2$. There are (up to) two stages of info acquisition:
 - Stage 1: source i with smaller noise ζ_i receives all attention
 - Stage 2: each source i receives fraction $\frac{\zeta_i}{\zeta_1 + \zeta_2}$
- Firms face a tradeoff between optimizing for long-run viewership and competing to be chosen in the short run
- Parameters that emphasize the long-run lead to higher eq ζ_i :
 - More patience (lower discount rate r)

Intuition

- Suppose $\phi_1 = \phi_2$. There are (up to) two stages of info acquisition:
 - Stage 1: source i with smaller noise ζ_i receives all attention
 - Stage 2: each source i receives fraction $\frac{\zeta_i}{\zeta_1 + \zeta_2}$
- Firms face a tradeoff between optimizing for long-run viewership and competing to be chosen in the short run
- Parameters that emphasize the long-run lead to higher eq ζ_i :
 - More patience (lower discount rate r)
 - Higher incentives for bias (larger λ and κ):
 - Polarized news sources live in symbiosis: provide complementary information, Stage 2 starts earlier.

Application 3: Attention Manipulation

Attention Grabbing

- Suppose a third party temporarily diverts the agent's attention towards source i .
 - ① Does this lead to a persistently higher amount of attention devoted to source i ?
 - ② What are the attention externalities on other sources?

Gossner, Steiner, and Stewart (2020)

- Gossner et al (2020) recently studied these questions in a model in which an agent sequentially learns about the quality of a number of goods.

Gossner, Steiner, and Stewart (2020)

- Gossner et al (2020) recently studied these questions in a model in which an agent sequentially learns about the quality of a number of goods.
- One of their main results resolves the two questions in the following way.
 - ① Does manipulation of attention towards i lead to a persistently higher amount of attention devoted to that source?
→ Yes, cumulated attention to that source is higher at every subsequent moment.
 - ② What are the attention externalities on other sources?
→ Cumulative attention paid to any other source is lower at every subsequent moment.

Our Frameworks are Non-Nested

Key assumption in GSS: attention strategy used by the agent satisfies a version of Independence of Irrelevant Alternatives (IIA):

Conditional on not focusing on the good to which attention is diverted, the agent's belief about that good does not affect the relative probabilities of focusing on the remaining goods.

Our Frameworks are Non-Nested

Key assumption in GSS: attention strategy used by the agent satisfies a version of Independence of Irrelevant Alternatives (IIA):

Conditional on not focusing on the good to which attention is diverted, the agent's belief about that good does not affect the relative probabilities of focusing on the remaining goods.

Our framework differs in a few key ways:

- agent learns about multiple attributes of an unknown (uni-dimensional) payoff-relevant state
- attribute values may be correlated

Outside of the special case of independence, the optimal attention allocation strategy generally fails IIA.

Applying our Characterization

Our characterization applies for any Σ satisfying one of the given conditions.

Thus can apply the characterization also **off path** given the posterior covariance matrix following manipulation.

Revisiting GSS's Findings

- 1 Does manipulation of attention towards i lead to a persistently higher amount of attention devoted to that source?

→ Yes, cumulated attention to that source is higher at every subsequent moment.

- 2 What are the attention externalities on other sources?

→ Cumulative attention paid to any other source is lower at every subsequent moment.

Revisiting GSS's Findings

- 1 Does manipulation of attention towards i lead to a persistently higher amount of attention devoted to that source?

→ Yes, cumulated attention to that source is higher at every subsequent moment.

HOLDS FOR ARBITRARY CORRELATION

- 2 What are the attention externalities on other sources?

→ Cumulative attention paid to any other source is lower at every subsequent moment.

Revisiting GSS's Findings

- 1 Does manipulation of attention towards i lead to a persistently higher amount of attention devoted to that source?

→ Yes, cumulated attention to that source is higher at every subsequent moment.

HOLDS FOR ARBITRARY CORRELATION

- 2 What are the attention externalities on other sources?

→ Cumulative attention paid to any other source is lower at every subsequent moment.

IN GENERAL CAN BREAK, BUT HOLDS UNDER AN ADDITIONAL CONDITION

Our Findings

Proposition

- 1 *For any Σ satisfying sufficient conditions, manipulation of attention towards source i leads to weakly higher cumulative attention at every subsequent time.*

Our Findings

Perpetual Substitutes: Σ^{-1} has negative off-diagonal entries, i.e. every pair of attributes has positive partial correlation.

Proposition

- 1 *For any Σ satisfying sufficient conditions, manipulation of attention towards source i leads to weakly higher cumulative attention at every subsequent time.*
- 2 *Suppose Perpetual Substitutes is satisfied.
Then manipulation of attention towards source i leads to weakly lower cumulative attention towards every other source at every moment of time.*

Summary of Applications

- **Application 1: Binary Choice**

- DM learns about unknown payoffs between two goods before making a choice
- use our characterization to generalize recent results from Fudenberg et al (2018)

- **Application 2: Biased News Sources**

- game between biased sources providing information over time
- use our characterization to solve for equilibrium

- **Application 3: Attention Manipulation**

- dynamic implications of a one-shot attention manipulation
- use our characterization to derive complementary results to Gossner et al (2020)

Conclusion

- We characterize optimal dynamic allocation of attention across multiple correlated information sources.
- Under weak conditions on the prior belief, the solution has a simple structure, is history-independent, and is robust across decision problems.
- Useful for applications!

Possible Extensions

Results hold also for:

- discrete model where agents allocate a fixed budget of precisions each period
- discrete model where agents choose a budget size of precisions each period (at some cost) and allocate it
- intertemporal decision problems (choose actions over time as well, receive payoff that depends on the sequence of actions)

Thank You!