Influence Campaigns

Evan Sadler

Columbia University

Paris School of Economics, October 8, 2018
Motivation

Huge investments to manipulate opinions and beliefs

• 2016 US elections (ongoing Russian interference...)
• Political advocacy (climate change)
• Public health (vaccines)
• Advertising

Some common features

• Expressed opinions and beliefs are coarse and diverse
• People are influenced by their friends
• Heterogeneous susceptibility to influence

High-level question: how to evaluate and design influence campaigns?
Motivation

Huge investments to manipulate opinions and beliefs

- 2016 US elections (ongoing Russian interference...)

Some common features

- Expressed opinions and beliefs are coarse and diverse
- People are influenced by their friends
- Heterogeneous susceptibility to influence

High-level question: how to evaluate and design influence campaigns?
Motivation

Huge investments to manipulate opinions and beliefs
- 2016 US elections (ongoing Russian interference...)
- Political advocacy (climate change)
Motivation

Huge investments to manipulate opinions and beliefs

- 2016 US elections (ongoing Russian interference...)
- Political advocacy (climate change)
- Public health (vaccines)
Motivation

Huge investments to manipulate opinions and beliefs
- 2016 US elections (ongoing Russian interference...)
- Political advocacy (climate change)
- Public health (vaccines)
- Advertising
Motivation

Huge investments to manipulate opinions and beliefs
• 2016 US elections (ongoing Russian interference...)
• Political advocacy (climate change)
• Public health (vaccines)
• Advertising

Some common features
• Expressed opinions and beliefs are coarse and diverse
• People are influenced by their friends
• Heterogeneous susceptibility to influence
Motivation

Huge investments to manipulate opinions and beliefs
- 2016 US elections (ongoing Russian interference...)
- Political advocacy (climate change)
- Public health (vaccines)
- Advertising

Some common features
- Expressed opinions and beliefs are coarse and diverse
- People are influenced by their friends
- Heterogeneous susceptibility to influence

High-level question: how to evaluate and design influence campaigns?
Learning models seem ill-suited for these settings.
Motivation, continued

Learning models seem ill-suited for these settings

Consensus is a robust feature of social learning models

- Both Bayesian and Non-Bayesian models...
Motivation, continued

Learning models seem ill-suited for these settings

Consensus is a robust feature of social learning models
  • Both Bayesian and Non-Bayesian models...

One-dimensional influence
  • In connected graph, I influence all others equally
A Different Approach

Agents in a network communicate coarse beliefs
A Different Approach

Agents in a network communicate coarse beliefs

Some are stubborn, some are open minded

- Beliefs evolve according to a Markov process
Agents in a network communicate coarse beliefs

Some are stubborn, some are open minded
- Beliefs evolve according to a Markov process

Stubborn agents exert influence
- Characterize influence of each stubborn agent on each open agent
A Different Approach

Agents in a network communicate coarse beliefs

Some are stubborn, some are open minded

• Beliefs evolve according to a Markov process

Stubborn agents exert influence

• Characterize influence of each stubborn agent on each open agent

Framework naturally describes echo chambers, can evaluate different interventions
Road Map

Basic framework and interpretation

Social learning questions: consensus, influence, aggregation

Random graphs and steady-state beliefs

Interventions: changing-minds, sowing doubt

Polarization as an objective
Related Work

Social Learning literature

- Bayesian (Acemoglu et al. 2011; Lobel and Sadler 2015, 2016)
- Non-Bayesian (Golub and Jackson 2010, 2012; Molavi et al. 2018)
- With misspecified models (Bohren and Hauser 2016, 2018)
Related Work

Social Learning literature

- Bayesian (Acemoglu et al. 2011; Lobel and Sadler 2015, 2016)
- Non-Bayesian (Golub and Jackson 2010, 2012; Molavi et al. 2018)
- With misspecified models (Bohren and Hauser 2016, 2018)

Voter models (Yildiz et al. 2013; Mossel and Tamuz 2014)
$n$ agents, unknown binary state $\theta \in \{0, 1\}$
Communicating Beliefs in a Network

$n$ agents, unknown binary state $\theta \in \{0, 1\}$

Three possible beliefs $\{\emptyset, 0, 1\}$
- Belief in states 0 or 1
- Uncertainty $\emptyset$
$n$ agents, unknown binary state $\theta \in \{0, 1\}$

Three possible beliefs $\{\emptyset, 0, 1\}$
- Belief in states 0 or 1
- Uncertainty $\emptyset$

Agents communicate claims about the world
- Agent $i$ believes $B \in \{\emptyset, 0, 1\}$, hears $P \in \{0, 1\}$, updates to $U_i(B, P) \in \{B, P\}$
Agents are linked in an undirected graph $G$

- Assume connected
The Communication Network

Agents are linked in an undirected graph $G$

- Assume connected

In each period, a directed edge $ij$ is selected uniformly at random

- Agent $i$ communicates her belief (if non-empty) to agent $j$
- Agent $j$ updates (if a claim is communicated)
The Communication Network

Agents are linked in an undirected graph $G$
- Assume connected

In each period, a directed edge $ij$ is selected uniformly at random
- Agent $i$ communicates her belief (if non-empty) to agent $j$
- Agent $j$ updates (if a claim is communicated)

Assume 1 is true, initial beliefs i.i.d with $\mathbb{P}(B = 1) = p \in \left(\frac{1}{2}, 1\right)$
Update Rules

Many possible update rules $U$, focus on three (mostly first two)

- **Open Rule:** $U(B, P) = P$
- **Stubborn Rule:** $U(B, P) = B$ whenever $B \in \{0, 1\}$
- **Skeptic Rule:** $U(\emptyset, P) = \emptyset$
Update Rules

Many possible update rules $U$, focus on three (mostly first two)

- Open Rule: $U(B, P) = P$
- Stubborn Rule: $U(B, P) = B$ whenever $B \in \{0, 1\}$
- Skeptic Rule: $U(\emptyset, P) = \emptyset$

Rules derived from an axiomatic foundation

- Weak credulity and label neutrality
- Order independence and label neutrality
Update Rules

Many possible update rules $U$, focus on three (mostly first two)

- Open Rule: $U(B, P) = P$
- Stubborn Rule: $U(B, P) = B$ whenever $B \in \{0, 1\}$
- Skeptic Rule: $U(\emptyset, P) = \emptyset$

Rules derived from an axiomatic foundation

- Weak credulity and label neutrality
- Order independence and label neutrality

Going forward, assume a mix of open and stubborn

- Skeptics enter when we talk about interventions
Interpreting Update Rules

Interpret as either opinion exchange or beliefs about some objective fact

- e.g. whether humans cause climate change, whether vaccines cause autism, whether some rumor is true
Interpreting Update Rules

Interpret as either opinion exchange or beliefs about some objective fact

• e.g. whether humans cause climate change, whether vaccines cause autism, whether some rumor is true

Coarse beliefs without direct verification
Interpreting Update Rules

Interpret as either opinion exchange or beliefs about some objective fact

- e.g. whether humans cause climate change, whether vaccines cause autism, whether some rumor is true

Coarse beliefs without direct verification

Stubborn agents could be partisans or experts
Interpreting Update Rules

Interpret as either opinion exchange or beliefs about some objective fact
• e.g. whether humans cause climate change, whether vaccines cause autism, whether some rumor is true

Coarse beliefs without direct verification

Stubborn agents could be partisans or experts

Open agents trust their neighbors
• Paul Grice and the “cooperative principle”
A Simple Example: Echo Chambers

Suppose the network consists of two cliques

• Clique 0 contains $m_0$ agents
• Clique 1 contains $m_1$ agents
• A single pair of agents is linked across the cliques
A Simple Example: Echo Chambers

Suppose the network consists of two cliques

- Clique 0 contains $m_0$ agents
- Clique 1 contains $m_1$ agents
- A single pair of agents is linked across the cliques

Beliefs evolve according to a finite Markov chain, unique stationary distribution

- Write $x_0, x_1$ for expected beliefs of typical agents
- Write $\bar{x}_0, \bar{x}_1$ for expected beliefs of the boundary agents
A Simple Example: Echo Chambers

Suppose the network consists of two cliques

• Clique 0 contains $m_0$ agents
• Clique 1 contains $m_1$ agents
• A single pair of agents is linked across the cliques

Beliefs evolve according to a finite Markov chain, unique stationary distribution

• Write $x_0, x_1$ for expected beliefs of typical agents
• Write $\overline{x}_0, \overline{x}_1$ for expected beliefs of the boundary agents

Clique 0 contains a single stubborn agent with belief 0, clique 1 contains a single stubborn agent with belief 1
• Stubborn agents exert influence
A Simple Example: Echo Chambers
A Simple Example: Echo Chambers

In this example, symmetry makes it easy to compute:

\[ x_0 = \frac{m_1}{m_0 m_1 + 2m_0 + 2m_1}, \quad x_1 = \frac{m_0 m_1 + m_0 + 2m_1}{m_0 m_1 + 2m_0 + 2m_1}, \]

\[ \bar{x}_0 = 2x_0, \quad \bar{x}_1 = 2x_1 - 1. \]
A Simple Example: Echo Chambers

In this example, symmetry makes it easy to compute:

\[ x_0 = \frac{m_1}{m_0 m_1 + 2m_0 + 2m_1}, \quad x_1 = \frac{m_0 m_1 + m_0 + 2m_1}{m_0 m_1 + 2m_0 + 2m_1}, \]

\[ \bar{x}_0 = 2x_0, \quad \bar{x}_1 = 2x_1 - 1. \]

Observations:
- Bigger cliques \( \implies \) more polarization
- Size of own clique limits influence of other clique
Initial Questions

Will agents reach a consensus?

- Usual answer: essentially always
- Usual answer: the most "central" agents
- Usual answer: yes, as long as no one is "too" influential

We're going to get very different answers
Initial Questions

Will agents reach a consensus?
- Usual answer: essentially always
Initial Questions

Will agents reach a consensus?
• Usual answer: essentially always

Who is most influential?
Initial Questions

Will agents reach a consensus?
• Usual answer: essentially always

Who is most influential?
• Usual answer: the most “central” agents
Initial Questions

Will agents reach a consensus?
• Usual answer: essentially always

Who is most influential?
• Usual answer: the most “central” agents

Will society aggregate information?
Initial Questions

Will agents reach a consensus?
• Usual answer: essentially always

Who is most influential?
• Usual answer: the most “central” agents

Will society aggregate information?
• Usual answer: yes, as long as no one is “too” influential
Initial Questions

Will agents reach a consensus?
  • Usual answer: essentially always

Who is most influential?
  • Usual answer: the most “central” agents

Will society aggregate information?
  • Usual answer: yes, as long as no one is “too” influential

We’re going to get very different answers
Network achieves *consensus* if at some point everyone has the same belief.
Disagreement

Network achieves consensus if at some point everyone has the same belief

**Theorem**

We get consensus w.p. 1 iff there is at most one stubborn agent. In a sequence of graphs \( \{G_n\} \), if the number of stubborn agents tends to infinity, the probability of reaching consensus tends to zero.
Disagreement

Network achieves consensus if at some point everyone has the same belief.

**Theorem**

We get consensus w.p. 1 iff there is at most one stubborn agent. In a sequence of graphs $\{G_n\}$, if the number of stubborn agents tends to infinity, the probability of reaching consensus tends to zero.

Proof: Trivial
Define *direct influence network* $\tilde{G}$:

- Let $d_i$ be degree of agent $i$, let $m$ be sum of all degrees
- Set of stubborn agents $S$
- If $i \in S$, take $\tilde{g}_{ii} = 1$, $\tilde{g}_{ij} = 0$ for $j \neq i$
- If $i \not\in S$, take $\tilde{g}_{ii} = \frac{m-d_i}{m}$, $\tilde{g}_{ij} = \frac{1}{m}$ if $j$ is a neighbor and 0 otherwise

Entry $\tilde{g}_{ij}$ gives probability that $i$ adopts $j$'s belief in any period.

Matrix powers $\tilde{G}^t$ describe probability of adopting particular beliefs at each period.

Limit as $t \to \infty$ describes steady-state beliefs.
Influence

Define *direct influence network* $\tilde{\mathcal{G}}$:

- Let $d_i$ be degree of agent $i$, let $m$ be sum of all degrees
- Set of stubborn agents $S$
- If $i \in S$, take $\tilde{g}_{ii} = 1$, $\tilde{g}_{ij} = 0$ for $j \neq i$
- If $i \notin S$, take $\tilde{g}_{ii} = \frac{m-d_i}{m}$, $\tilde{g}_{ij} = \frac{1}{m}$ if $j$ is a neighbor and 0 otherwise

Entry $\tilde{g}_{ij}$ gives probability that $i$ adopts $j$’s belief in any period
Define *direct influence network* $\tilde{G}$:

- Let $d_i$ be degree of agent $i$, let $m$ be sum of all degrees
- Set of stubborn agents $S$
- If $i \in S$, take $\tilde{g}_{ii} = 1$, $\tilde{g}_{ij} = 0$ for $j \neq i$
- If $i \notin S$, take $\tilde{g}_{ii} = \frac{m-d_i}{m}$, $\tilde{g}_{ij} = \frac{1}{m}$ if $j$ is a neighbor and 0 otherwise

Entry $\tilde{g}_{ij}$ gives probability that $i$ adopts $j$’s belief in any period

Matrix powers $\tilde{G}^t$ describe probability of adopting particular beliefs at each period

- Limit as $t \to \infty$ describes steady-state beliefs
Influence, continued

\( \tilde{G} \) is a row-stochastic matrix
- Corresponds to transitions for a finite state Markov chain
- Stubborn agents correspond to absorbing states
\( \tilde{G} \) is a row-stochastic matrix

- Corresponds to transitions for a finite state Markov chain
- Stubborn agents correspond to absorbing states

**Theorem**

There exists a unique \( M = \lim_{t \to \infty} \tilde{G}^t \) satisfying

\[
M = \sum_{i \in S} \nu^{(i)} e'_i
\]

where \( e_i \) is a unit vector and the \( \{ \nu^{(i)} \} \) are linearly independent right eigenvectors of \( \tilde{G} \) with eigenvalue 1 such that \( \nu^{(i)}_i = 1 \) and \( \nu^{(i)}_j = 0 \) for \( j \neq i, j \in S \).
Influence, continued

If $i \in S$, the $i$th column of $M$ is the eigenvector $\nu^{(i)}$

- Entries of $\nu^{(i)}$ describe influence of $i$ on each other agent $j$
If $i \in S$, the $i$th column of $M$ is the eigenvector $\nu^{(i)}$

- Entries of $\nu^{(i)}$ describe influence of $i$ on each other agent $j$

Influence depends both on network and on who else is stubborn
If \( i \in S \), the \( i \)th column of \( M \) is the eigenvector \( \nu^{(i)} \)

- Entries of \( \nu^{(i)} \) describe influence of \( i \) on each other agent \( j \)

Influence depends both on network and on who else is stubborn

Additional intuition: relationship to random walks

- Influential stubborn agents have lots of short paths to others
- Paths blocked by other stubborn agents don’t count
Computing Influence

Steady-state belief $x_i$, probability that $i$ believes 1
Computing Influence

Steady-state belief $x_i$, probability that $i$ believes 1

Submatrices $\tilde{G}_{-S,-S} (n - |S| \times n - |S|)$, $\tilde{G}_{-S,S} (n - |S| \times |S|)$
Computing Influence

Steady-state belief $x_i$, probability that $i$ believes 1

Submatrices $\tilde{G}_{-S,-S} (n - |S| \times n - |S|)$, $\tilde{G}_{-S,S} (n - |S| \times |S|)$

Steady-state condition:

$$x_{-S} = \tilde{G}_{-S,S} x + \tilde{G}_{-S,-S} x_{-S},$$

$$\implies x_{-S} = \left( I - \tilde{G}_{-S,-S} \right)^{-1} \tilde{G}_{-S,S} x_S$$
No Aggregation

Corollary

For each agent $i$ and each time period $t$, the probability that $i$ holds correct beliefs is $p$. 
Corollary

For each agent $i$ and each time period $t$, the probability that $i$ holds correct beliefs is $p$.

Probability of correct belief is just a linear combination of initial probabilities

- Without some bias towards truth in the update process, no learning is possible
Towards Intervention Design

Practical issues:
• Network data is expensive and imperfect
• These matrices can be large, computationally intensive
Towards Intervention Design

Practical issues:
- Network data is expensive and imperfect
- These matrices can be large, computationally intensive

My approach: random graphs, large network limits
Towards Intervention Design

Practical issues:
• Network data is expensive and imperfect
• These matrices can be large, computationally intensive

My approach: random graphs, large network limits

The stochastic block model:
• Finite collection of types $\Theta$
• Types $\theta$ and $\theta'$ are linked with independent probability $p_{\theta,\theta'}$
• Take limit as $n \to \infty$, fraction of type $\theta$ converges to $q_\theta$
Influence in Random Graphs

- Define $d_\theta \equiv \sum_{\theta'} q_{\theta'} p_{\theta,\theta'}$
- Fraction $s_\theta$ of type $\theta$ is stubborn
Influence in Random Graphs

• Define $d_\theta \equiv \sum_{\theta'} q_{\theta'} p_{\theta,\theta'}$

• Fraction $s_\theta$ of type $\theta$ is stubborn

Weights

$$w^{(S)}_{\theta,\theta'} = \frac{s_{\theta'} q_{\theta'} p_{\theta,\theta'}}{d_\theta}, \quad w^{(-S)}_{\theta,\theta'} = \frac{(1 - s_{\theta'}) q_{\theta'} p_{\theta,\theta'}}{d_\theta}$$

describe fraction of neighbors who are of each type and stubborn (or not)
As network gets large, suppose fraction of stubborn agents of type $\theta$ who believe 1 converges to $x_{\theta}^{(S)}$.
As network gets large, suppose fraction of stubborn agents of type $\theta$ who believe 1 converges to $x^{(S)}_{\theta}$

Steady-state beliefs for open type $\theta$ agents converge to $x^{(-S)}_{\theta}$, where

$$x^{(-S)} = \left( I - W^{(-S)} \right)^{-1} W^{(S)} x^{(S)}$$
Influence in Random Graphs

As network gets large, suppose fraction of stubborn agents of type $\theta$ who believe 1 converges to $x^{(S)}_{\theta}$

Steady-state beliefs for open type $\theta$ agents converge to $x^{(-S)}_{\theta}$, where

$$x^{(-S)} = (I - W^{(-S)})^{-1} W^{(S)} x^{(S)}$$

Matrix $\hat{R} = (I - W^{(-S)})^{-1} W^{(S)}$ gives influence of each type of stubborn agent on each type of open agent.
Influence in Random Graphs

As network gets large, suppose fraction of stubborn agents of type $\theta$ who believe 1 converges to $x_{\theta}^{(S)}$.

Steady-state beliefs for open type $\theta$ agents converge to $x_{\theta}^{(-S)}$, where

$$x^{(-S)} = \left(I - W^{(-S)}\right)^{-1} W^{(S)} x^{(S)}$$

Matrix $\hat{R} = \left(I - W^{(-S)}\right)^{-1} W^{(S)}$ gives influence of each type of stubborn agent on each type of open agent.

Issue: influence of types, not of individuals.
Normalized Influence

From matrix $\hat{R}$, define normalized matrix $R$ with entries

$$r_{\theta, \theta'} = \hat{r}_{\theta, \theta'} \frac{(1 - s_{\theta}) q_{\theta}}{s_{\theta'} q_{\theta'}}$$

Normalization gives a measure of how much a stubborn type $\theta'$ individual influences an open type $\theta$ individual.

Theorem

As $n \to \infty$, the average influence of a type $\theta'$ stubborn agent on a type $\theta$ open agent converges to $r_{\theta, \theta'}$. 

Evan Sadler
Influence Campaigns
24/42
Normalized Influence

From matrix $\hat{R}$, define normalized matrix $R$ with entries

$$r_{\theta,\theta'} = \hat{r}_{\theta,\theta'} \frac{(1 - s_\theta)q_\theta}{s_{\theta'}q_{\theta'}}$$

Normalization gives a measure of how much a stubborn type $\theta'$ individual influences an open type $\theta$ individual.
Normalized Influence

From matrix $\hat{R}$, define normalized matrix $R$ with entries

$$r_{\theta, \theta'} = \hat{r}_{\theta, \theta'} \frac{(1 - s_\theta) q_\theta}{s_{\theta'} q_{\theta'}}$$

Normalization gives a measure of how much a stubborn type $\theta'$ individual influences an open type $\theta$ individual.

**Theorem**

As $n \to \infty$, the average influence of a type $\theta'$ stubborn agent on a type $\theta$ open agent converges to $r_{\theta, \theta'}$
Interventions

Consider two interventions:

• Changing minds
• Sowing doubt
Interventions

Consider two interventions:

- Changing minds
- Sowing doubt

Changing minds is easiest to assess: if I convert a type $\theta$ stubborn agent, then column $\theta$ of $R$ describes the effect

- Note heterogeneous impact across other types
Interventions

Consider two interventions:

• Changing minds
• Sowing doubt

Changing minds is easiest to assess: if I convert a type $\theta$ stubborn agent, then column $\theta$ of $R$ describes the effect

• Note heterogeneous impact across other types

Effect of sowing doubt depends on indirect influence of open agents

• Depends on who influences an open agent, proportional to normalized columns of $W^{(-S)}$
Two Groups with Homophily

To make things more concrete, let’s work through an easy example

• There are two types, 1 and 2, equally prevalent
• Fractions $s_1, s_2$ are stubborn
• Link to own type w.p. $p_s$, to other type w.p. $p_d$
Two Groups with Homophily

To make things more concrete, let’s work through an easy example

- There are two types, 1 and 2, equally prevalent
- Fractions $s_1, s_2$ are stubborn
- Link to own type w.p. $p_s$, to other type w.p. $p_d$

First, look at influence as we vary:

- Homophily
- Stubbornness
The Role of Homophily

- Influence on Own Type
- Influence on Other Type

Fraction of Own-Type Neighbors
Asymmetric Stubbornness

Influence on Own Type

Influence on Other Type

Own Fraction of Stubborn Agents
Lessons on Influence

Increasing homophily trades off influence on own type against influence on other type

- Intuitive result: to affect views of one group, target influencers in that group
Lessons on Influence

Increasing homophily trades off influence on own type against influence on other type

- Intuitive result: to affect views of one group, target influencers in that group

Formal argument why even limited exposure to other group has big impact on polarization
Lessons on Influence

Increasing homophily trades off influence on own type against influence on other type

- Intuitive result: to affect views of one group, target influencers in that group

Formal argument why even limited exposure to other group has big impact on polarization

More individual influence if own group is less stubborn

- Total influence of type increases in stubbornness
- Individual members cannibalize others’ influence
Sowing Doubt

- Influence Campaigns

Influence of Own Type

Influence of Other Type

Fraction of Skeptics in Other Type
Sowing Doubt

Sowing doubt in one group increases influence of the other

- Driven by echo chamber effect, fewer people reinforce influence by repeating claims
Sowing Doubt

Sowing doubt in one group increases influence of the other

- Driven by echo chamber effect, fewer people reinforce influence by repeating claims

Increasing returns to sowing doubt
Sowing Doubt

Sowing doubt in one group increases influence of the other

- Driven by echo chamber effect, fewer people reinforce influence by repeating claims

Increasing returns to sowing doubt

Absent from the graph: influence on other group goes down

- Fewer agents willing to adopt a certain belief
- Relative influence moves towards the non-skeptical group
Polarization and Variance

Suppose the population will take a collective decision

- Depends on (random) beliefs at a particular moment
Polarization and Variance

Suppose the population will take a collective decision

- Depends on (random) beliefs at a particular moment

If I want to affect the outcome, I care about expectation \textit{and} variance of beliefs
Polarization and Variance

Suppose the population will take a collective decision

- Depends on (random) beliefs at a particular moment

If I want to affect the outcome, I care about expectation and variance of beliefs

In stochastic block model, individual beliefs are asymptotically independent

- Variance of number of ones is approximately

\[
\sum_{i} \mathbb{E}[x_i](1 - \mathbb{E}[x_i])
\]
Polarization and Variance

Suppose the population will take a collective decision

• Depends on (random) beliefs at a particular moment

If I want to affect the outcome, I care about expectation \textit{and} variance of beliefs

In stochastic block model, individual beliefs are asymptotically independent

• Variance of number of ones is approximately

\[ \sum_i \mathbb{E}[x_i](1 - \mathbb{E}[x_i]) \]

More polarization \iff lower variance
Final Remarks

Social learning model with different answers
  • No consensus
  • Multifaceted influence
  • No aggregation

Key ingredients: coarse beliefs, reliance on unverified reports

Practical tool to study interventions
  • Homophily and echo chambers
  • Sowing doubt

Polarization and risk aversion
States, Propositions, Beliefs

Finite set of states of the world \( S \), true state \( s^* \)
States, Propositions, Beliefs

Finite set of states of the world $S$, true state $s^*$

A proposition $P \subseteq S$ is a subset of states

- True if $s^* \in P$
- Propositions $P$ and $P'$ contradict each other if $P \cap P' = \emptyset$
States, Propositions, Beliefs

Finite set of states of the world $S$, true state $s^*$

A proposition $P \subseteq S$ is a subset of states
- True if $s^* \in P$
- Propositions $P$ and $P'$ **contradict** each other if $P \cap P' = \emptyset$

A set of beliefs $B$ is a collection of propositions $\{P_i\}_{i \in I}$
- **Consistent** if $P(B) \equiv \bigcap_{i \in I} P_i \neq \emptyset$
- Write $\mathcal{B}$ for set of all consistent beliefs
States, Propositions, Beliefs

Finite set of states of the world $S$, true state $s^*$

A proposition $P \subseteq S$ is a subset of states
- True if $s^* \in P$
- Propositions $P$ and $P'$ contradict each other if $P \cap P' = \emptyset$

A set of beliefs $B$ is a collection of propositions $\{P_i\}_{i \in I}$
- Consistent if $P(B) \equiv \bigcap_{i \in I} P_i \neq \emptyset$
- Write $B$ for set of all consistent beliefs

Say $B$ and $B'$ are equivalent, written $B \cong B'$, if $P(B) = P(B')$
Learning and Updating

Encounter propositions sequentially, must decide what to believe
Learning and Updating

Encounter propositions sequentially, must decide what to believe

At each instant, agent must hold consistent beliefs
Learning and Updating

Encounter propositions sequentially, must decide what to believe

At each instant, agent must hold consistent beliefs

Update rule $U(B, P) : \mathcal{B} \times 2^S \rightarrow \mathcal{B}$

- $U(B, P) \subseteq B \cup \{P\}$
Learning and Updating

Encounter propositions sequentially, must decide what to believe

At each instant, agent must hold consistent beliefs

Update rule $U(B, P) : \mathcal{B} \times 2^S \rightarrow \mathcal{B}$

- $U(B, P) \subseteq B \cup \{P\}$

Implicit assumption: updating depends only on current beliefs

- Order in which I added propositions does not affect updates
- Propositions I earlier rejected as false do not affect updates
Learning and Updating

Encounter propositions sequentially, must decide what to believe

At each instant, agent must hold consistent beliefs

Update rule \( U(B, P) : B \times 2^S \rightarrow B \)

- \( U(B, P) \subseteq B \cup \{P\} \)

Implicit assumption: updating depends only on current beliefs

- Order in which I added propositions does not affect updates
- Propositions I earlier rejected as false do not affect updates

Write \( L \) for an ordered list of propositions, \( U(B, L) \) for beliefs after processing a list, \( U(L) \equiv U(\emptyset, L) \)
Examples

Stubborn updating:
- Reject any new proposition that conflicts with current beliefs
- Order of propositions in a list $L$ clearly matters
Examples

Stubborn updating:
- Reject any new proposition that conflicts with current beliefs
- Order of propositions in a list $L$ clearly matters

Variant: Stubborn with preference for certainty (SPC)
Examples

Stubborn updating:
• Reject any new proposition that conflicts with current beliefs
• Order of propositions in a list $L$ clearly matters

Variant: Stubborn with preference for certainty (SPC)

Wishful thinking (WT) update rule:
• Start with a preference relation $\succ$ on states
• When updating, retain only propositions that allow you to entertain most preferred state
Examples

Stubborn updating:
• Reject any new proposition that conflicts with current beliefs
• Order of propositions in a list $L$ clearly matters

Variant: Stubborn with preference for certainty (SPC)

Wishful thinking (WT) update rule:
• Start with a preference relation $\succ$ on states
• When updating, retain only propositions that allow you to entertain most preferred state

Skeptic rule: reject all propositions
Learning Axioms

Is there a “good” update rule?

- I take an axiomatic approach

Axiom 0 (Stability): If $P \in B$, then $U(B, P) = B$.

- Tell me what I already believe, and my beliefs don’t change

Axiom 0’ (Own-Frame Independence): if $B \sim B'$, then $U(B, P) \sim U(B', P)$ for any $P$.
Learning Axioms

Is there a “good” update rule?

• I take an axiomatic approach

Axiom 0 (Stability): If $P \in B$, then $U(B, P) = B$

• Tell me what I already believe, and my beliefs don’t change
Learning Axioms

Is there a “good” update rule?
• I take an axiomatic approach

Axiom 0 (Stability): If $P \in B$, then $U(B, P) = B$
• Tell me what I already believe, and my beliefs don’t change

Axiom 0’ (Own-Frame Independence): if $B \equiv B'$, then $U(B, P) \equiv U(B', P)$ for any $P$
Learning Axioms

Is there a “good” update rule?

• I take an axiomatic approach

Axiom 0 (Stability): If $P \in B$, then $U(B, P) = B$

• Tell me what I already believe, and my beliefs don’t change

Axiom 0’ (Own-Frame Independence): if $B \cong B'$, then $U(B, P) \cong U(B', P)$ for any $P$

Divide axioms into two categories:

• Willingness-to-learn
• Non-manipulability
Willingness-to-Learn Axioms

Openness: For any initial beliefs $B$ and any state $s$, there exists a list of propositions $L$ such that

$$P(U(B, L)) = \{s\}$$
Willingness-to-Learn Axioms

Openness: For any initial beliefs $B$ and any state $s$, there exists a list of propositions $L$ such that

$$P(U(B, L)) = \{s\}$$

Credulity: For any $B$ and $P$, there is no consistent belief $B' \subseteq B \cup P$ such that $U(B, P)$ is a strict subset of $B'$
Willingness-to-Learn Axioms

Openness: For any initial beliefs $B$ and any state $s$, there exists a list of propositions $L$ such that

$$P(U(B, L)) = \{s\}$$

Credulity: For any $B$ and $P$, there is no consistent belief $B' \subseteq B \cup P$ such that $U(B, P)$ is a strict subset of $B'$

Weak Credulity: For any non-empty $L$, the beliefs $U(L)$ are non-empty
Non-Manipulability Axioms

Label neutrality: permuting state labels does not change the update rule. More formally, letting $\pi$ denote a permutation on the state space $S$, we have

$$U(B, P)_\pi = U(B_\pi, P_\pi)$$
Non-Manipulability Axioms

Label neutrality: permuting state labels does not change the update rule. More formally, letting $\pi$ denote a permutation on the state space $S$, we have

$$U(B, P)_\pi = U(B_\pi, P_\pi)$$

Order independence: permuting propositions in a list does not change beliefs. For all $L$, we have

$$U(L) = U(\pi(L))$$
Non-Manipulability Axioms

Label neutrality: permuting state labels does not change the update rule. More formally, letting $\pi$ denote a permutation on the state space $S$, we have

$$U(B, P)_{\pi} = U(B_\pi, P_\pi)$$

Order independence: permuting propositions in a list does not change beliefs. For all $L$, we have

$$U(L) = U(\pi(L))$$

Frame independence: for any consistent $L$, we have

$$P(U(L_1 + L + L_2)) = P(L_1 + P(L) + L_2)$$
Theorem

Suppose $S$ contains at least 5 states.

- There is no update rule satisfying openness and order independence
- There is no update rule satisfying credulity and order independence
- There is no update rule satisfying openness and frame independence
- There is no update rule satisfying credulity and frame independence
Proof of first part:

• Suppose $a$ and $b$ are distinct states, let $\overline{L}$ be a list of all propositions

• By openness, there exists $L_a$ and $L_b$ such that $P(U(\overline{L}, L_a)) = \{a\}$ and $P(U(\overline{L}, L_b)) = \{b\}$

• By order independence, we can reorder the list $\overline{L} + L_a$ so that duplicates appear consecutively, and propositions remain in the same order as in $\overline{L}$. Stability then implies $U(\overline{L}, L_a) = U(\overline{L})$

• Likewise, $U(\overline{L}, L_b) = U(\overline{L})$, which contradicts $a \neq b$

Other parts follow from similar constructions
Theorem

- If $U$ satisfies label neutrality, order independence, and frame independence, then $U$ is the skeptic rule.
- If $U$ satisfies order independence, OFI, and weak credulity, then $U$ is a wishful thinking rule.
- If $U$ satisfies credulity, OFI, and label neutrality, then $U$ is stubborn, except possibly for propositions containing only one state.
Representation Theorem

Theorem

- If $U$ satisfies label neutrality, order independence, and frame independence, then $U$ is the skeptic rule.
- If $U$ satisfies order independence, OFI, and weak credulity, then $U$ is a wishful thinking rule.
- If $U$ satisfies credulity, OFI, and label neutrality, then $U$ is stubborn, except possibly for propositions containing only one state.

Inherent tradeoffs
Representation Theorem

Theorem

- If $U$ satisfies label neutrality, order independence, and frame independence, then $U$ is the skeptic rule.
- If $U$ satisfies order independence, OFI, and weak credulity, then $U$ is a wishful thinking rule.
- If $U$ satisfies credulity, OFI, and label neutrality, then $U$ is stubborn, except possibly for propositions containing only one state.

Inherent tradeoffs

A case for stubborn rules: prioritize the cooperative principle.