Imperial Workshop on Intelligent Communications
19 June 2023, Imperial College London

Social Learning

Belief Formation and Diffusion Over Graphs

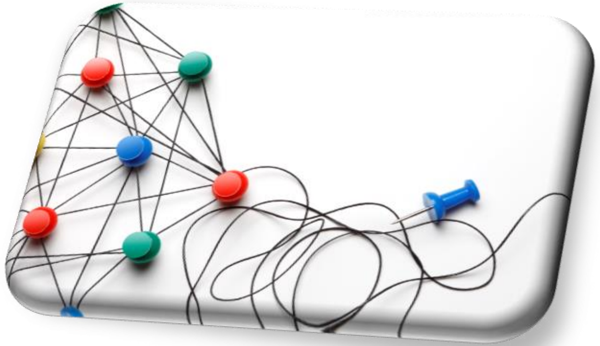
Vincenzo Matta

University of Salerno

What is Social Learning?

- **Humans** form their opinions via repeated interactions (even virtually over social networks)
- Nature provides splendid examples of cooperative learning in the form of **biological networks**

- Useful models across several disciplines: **Cognitive Sciences** (e.g., Psychology)**, Social Sciences** (e.g., Economics), **Statistics**,…



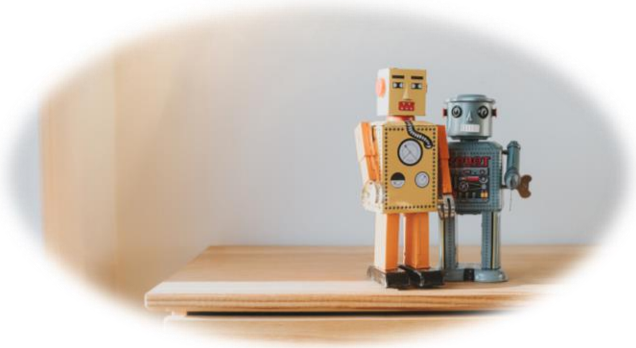
A Virtuous Circle

Man-engineered systems for **multi-agent decision-making**

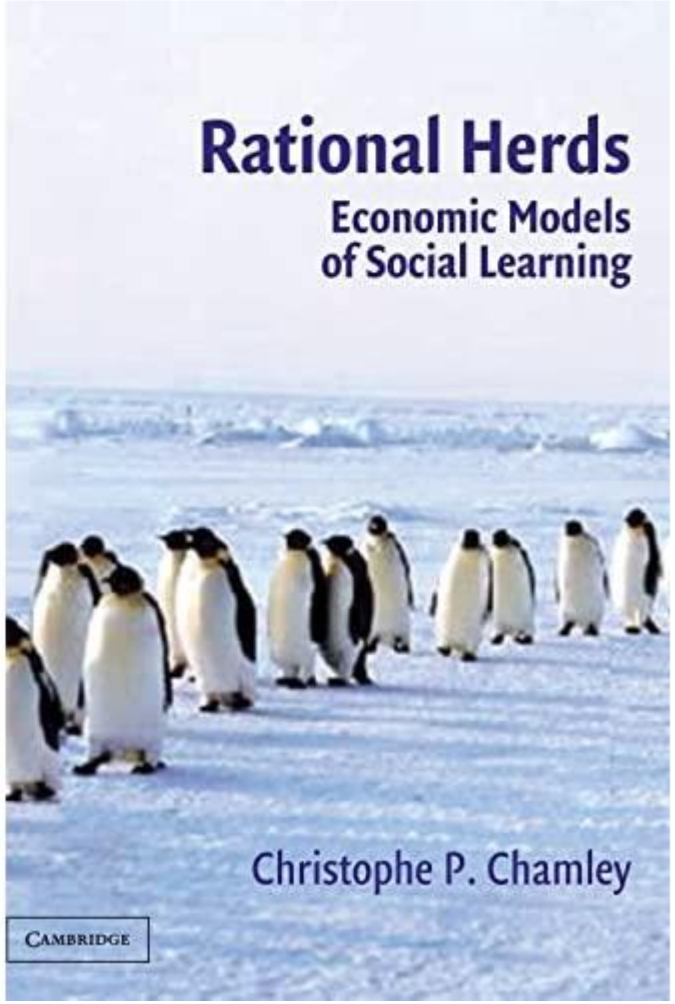
(IoT networks, mobile phones, robotic swarms,…)

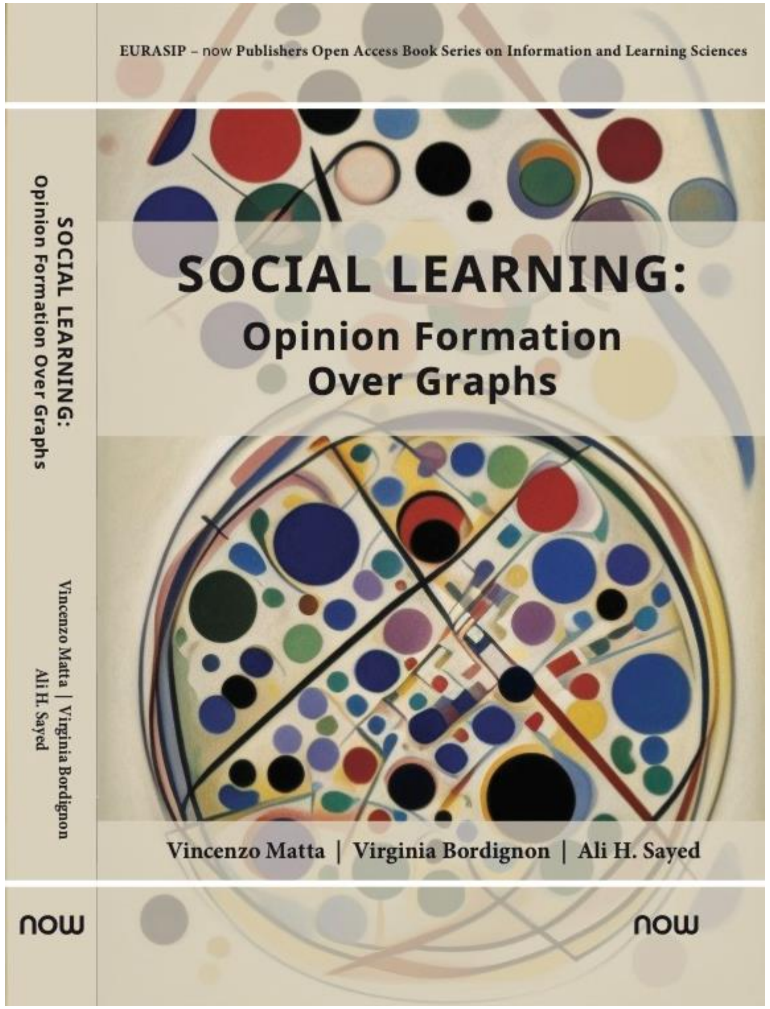
Nature

Engineering



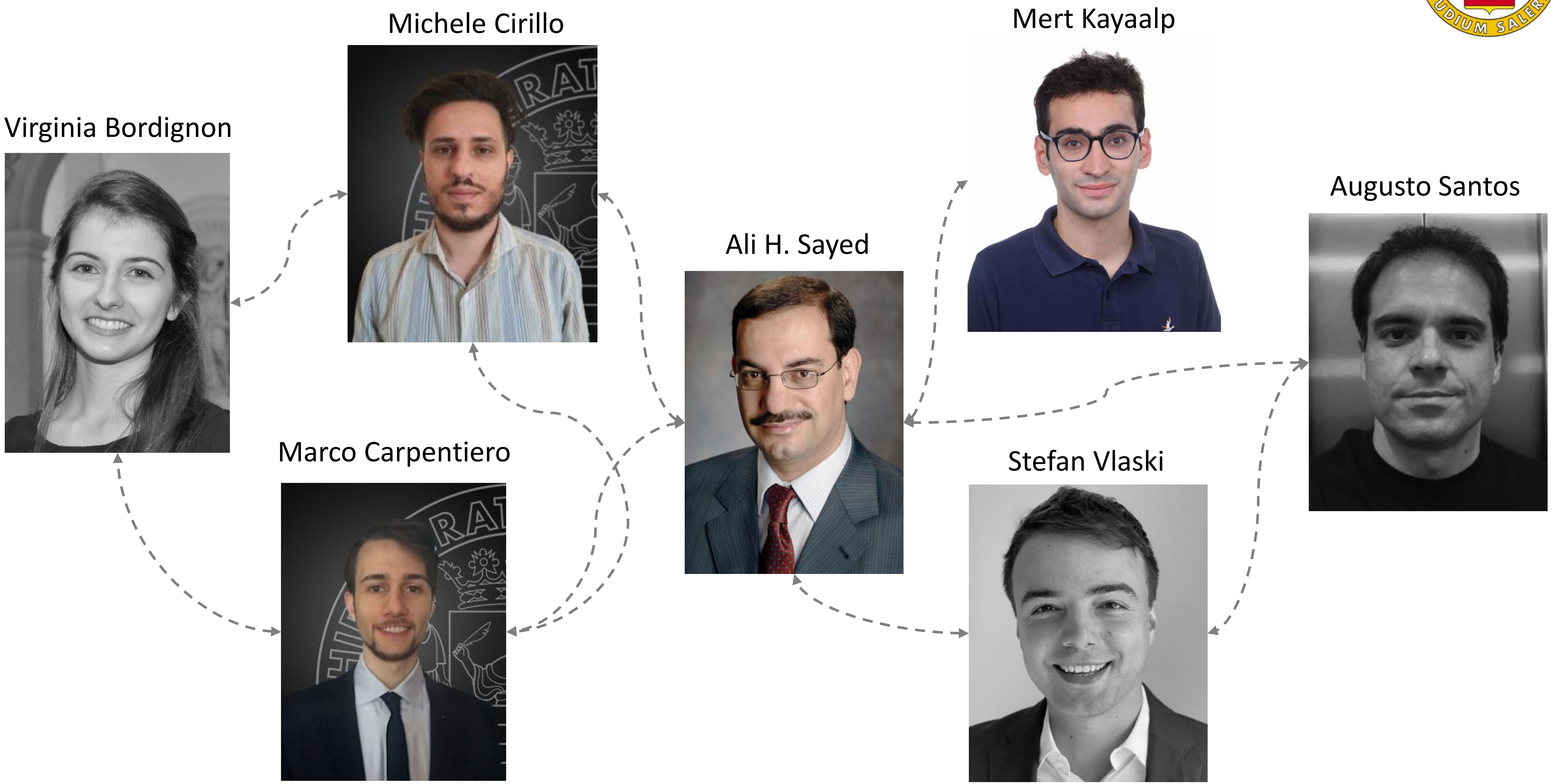
Useful References





Many other references focusing on different perspectives, e.g., psychological, behavioral, or biological aspects

Our Social Research Network



Outline

- **Part I. Traditional (single-agent) belief formation**

- **Part II. Social learning: Belief formation over graphs**
  - Agreement
  - Discord, influencers vs. influenced agents, fake news,…

- **Part III. Recent trends in social learning**
  - Adaptive social learning
  - Social learning with partial information
  - Social machine learning

Part I

Traditional (Single-Agent) Belief Formation

Hypotheses and Data

$\Theta = \{\theta_1, \theta_2, \ldots, \theta_M\}$ set of M hypotheses

which team wins?

injuries, last results,...

what weather tomorrow?

temperature, humidity,...

$\cdots$

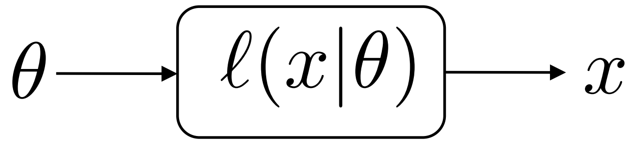
which is the best road?

traffic, context information,...

Data

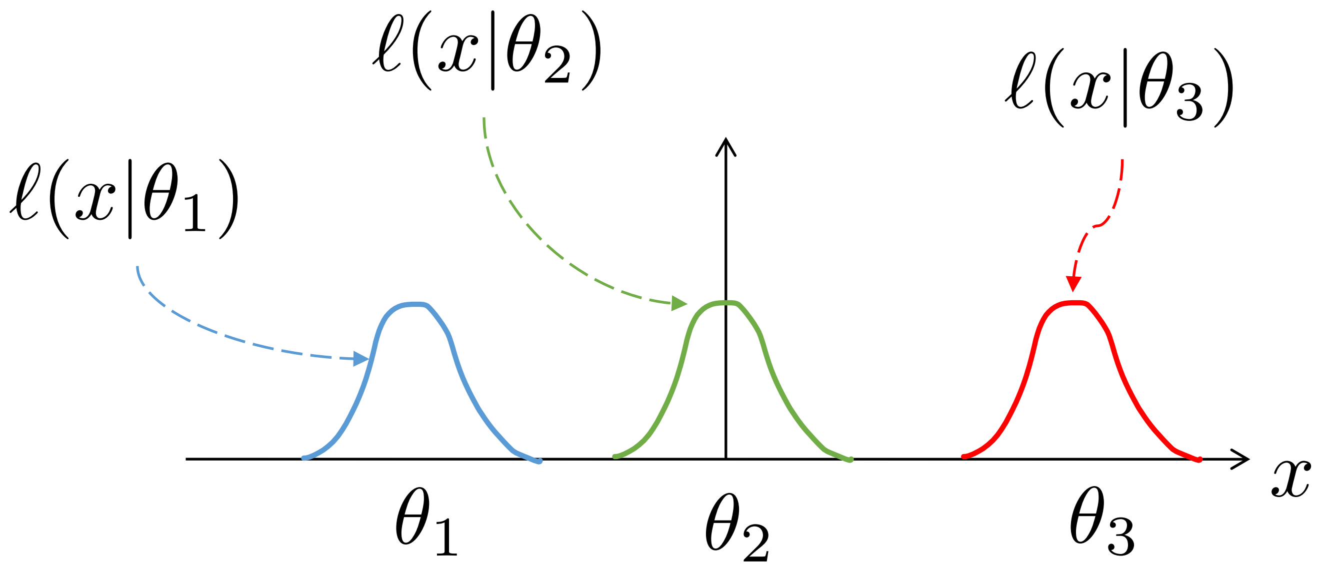
Likelihood

The likelihood encodes the **probabilistic mechanism** connecting the data to the hypothesis
e.g., which are typical values of humidity if it rains?



$$x \in \mathcal{X}$$

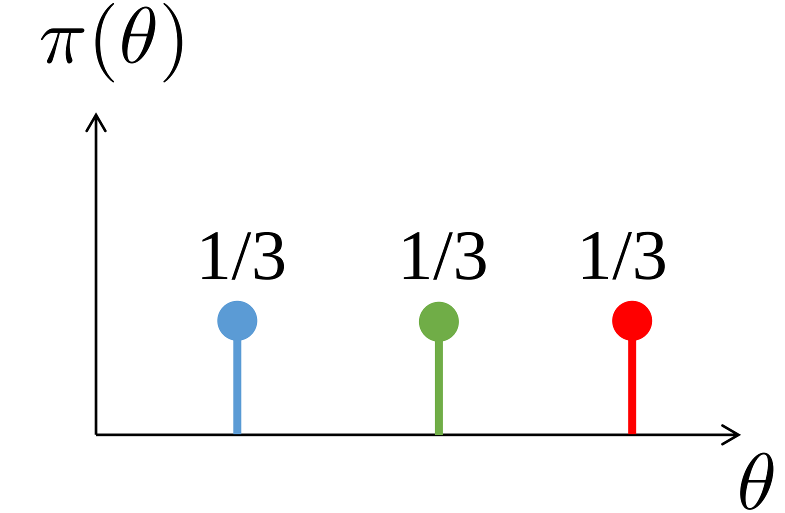
scalar / vector
continuous (pdf) / categorical (pmf) / mixed



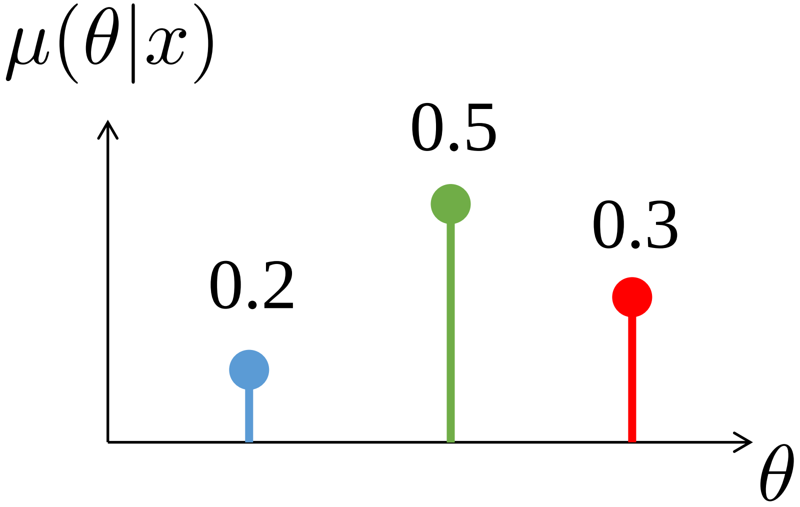
Belief

We assign probability scores to the hypotheses

prior belief (no data!)

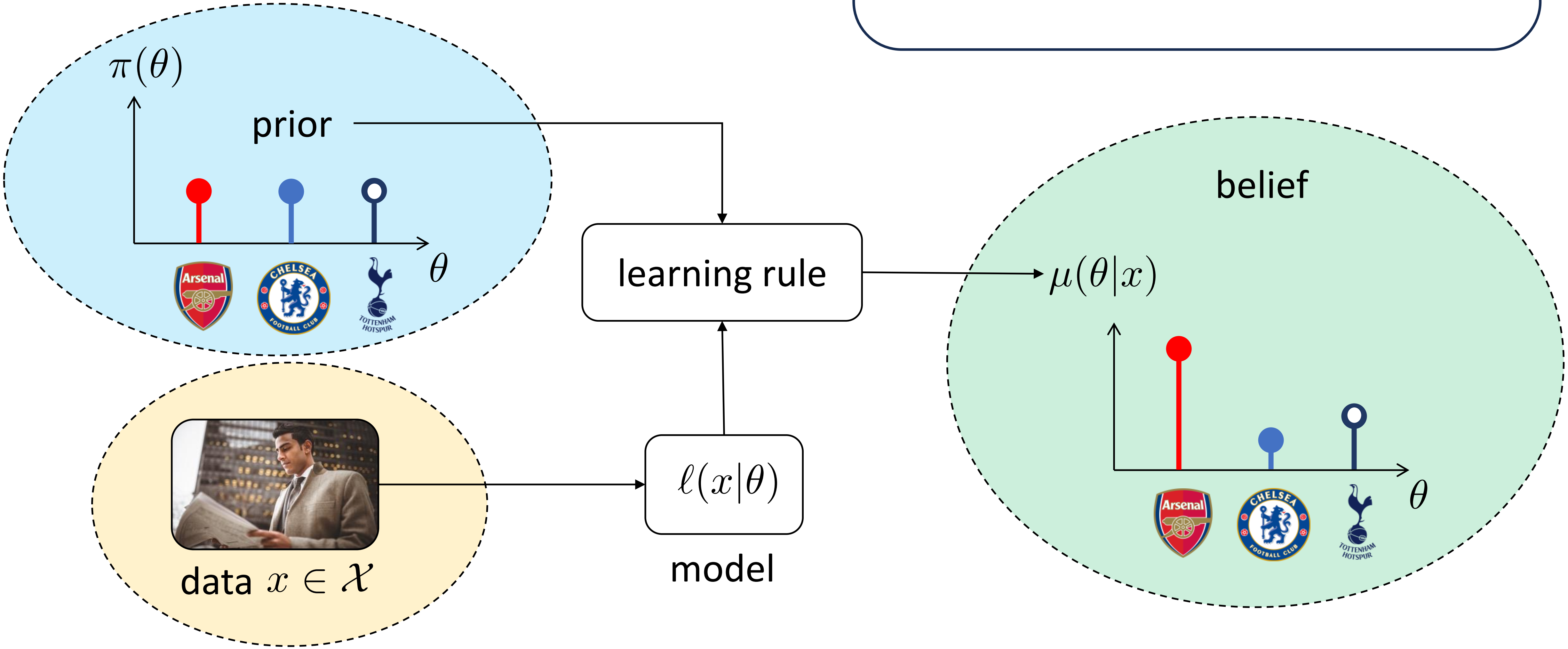


posterior belief **given the observed data**



Example

probability that a team has the highest rank at the end of the season



Bayes' Rule

build the posterior from the data

$$\mu(\theta|x) = \frac{\pi(\theta)\ell(x|\theta)}{m(x)}$$

Thomas Bayes
(1760-1761)

$$m(x) = \sum_{\theta \in \Theta} \pi(\theta)\ell(x|\theta)$$

marginal distribution of the data

$$\mu(\theta|x) \propto \pi(\theta)\ell(x|\theta)$$

hiding the
normalization factor

- One pillar of Probability Theory
- Optimal from an epistemological perspective
- Optimal from an information-theoretic perspective

  (see also the free-energy principle and variational Bayesian inference)
- Model for cognition: Bayesian brain [FristonKilnerHarrison2006]

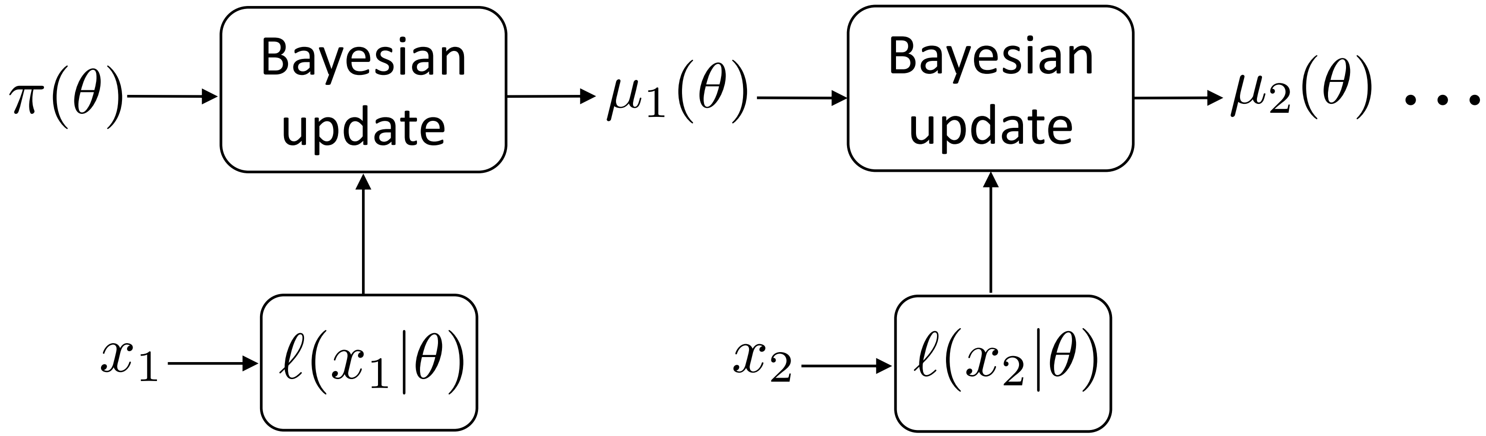
Sequential Bayesian Updates

$$\mu_t(\theta) \triangleq \mu(\theta | x_1, x_2, \ldots, x_t)$$

streaming data (iid)

All the **necessary knowledge is stored in the last belief**
The last belief becomes the **prior for the subsequent step**

$$\mu_t(\theta) \propto \mu_{t-1}(\theta)\ell(x_t|\theta)$$



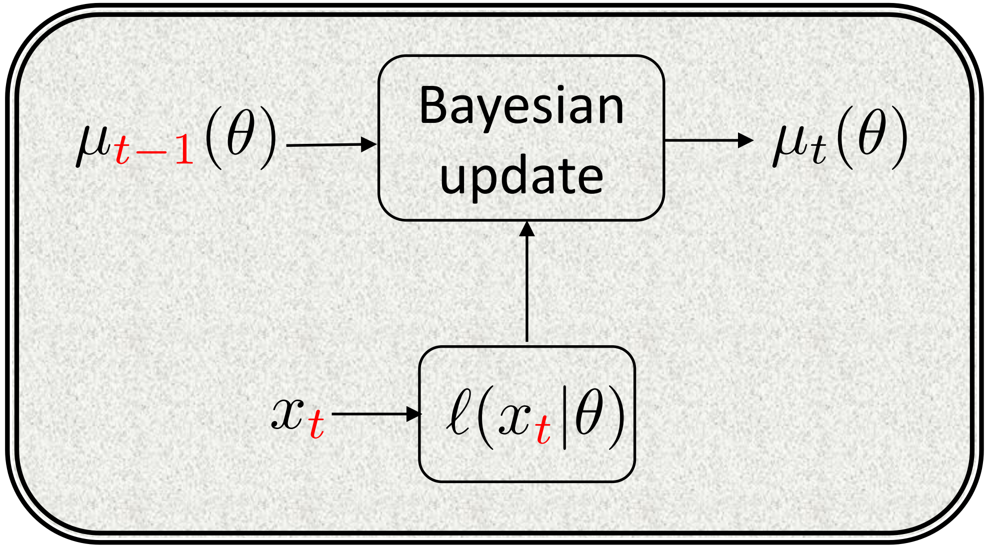
Sequential Bayesian Updates

$$\mu_t(\theta) \triangleq \mu(\theta | x_1, x_2, \ldots, x_t)$$

streaming data (iid)

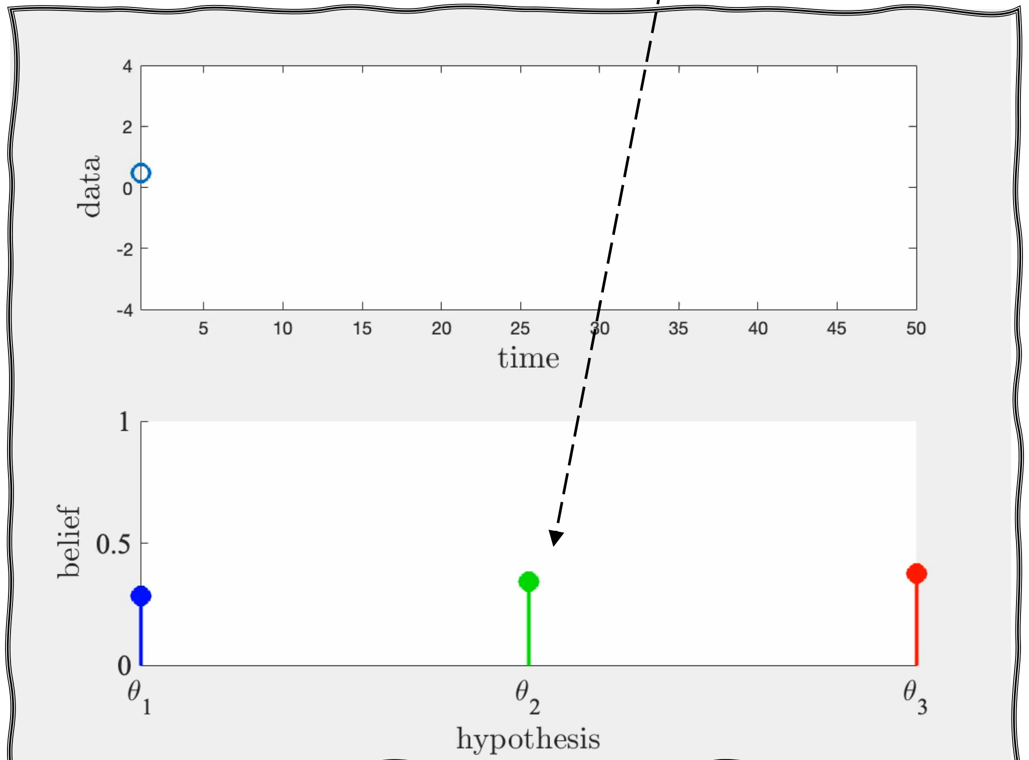
All the **necessary knowledge is stored in the last belief**
The last belief becomes the **prior for the subsequent step**

$$\mu_t(\theta) \propto \mu_{t-1}(\theta)\ell(x_t|\theta)$$



Bayesian Learning at Work

data generated according to hypothesis θ_2



Convergence of Bayesian Learning

$$D(f||\ell_\theta) \triangleq \mathbb{E}_f\left[\log \frac{f(\boldsymbol{x})}{\ell(\boldsymbol{x}|\theta)}\right]$$

The Kullback-Leibler divergence quantifies the discrepancy between two probability measures

Andrej Nikolaevič Kolmogorov
(1903-1987)

data generated according to distribution f

$$\theta^\star = \arg\min_{\theta\in\Theta} D(f||\ell_\theta)$$ [Berk1966]

Convergence to the likelihood featuring the highest match with the true distribution

$$\boldsymbol{\mu}_t(\theta^\star) \xrightarrow{t\to\infty} 1 \quad \text{almost surely}$$

Part II

Social Learning:
Belief Formation Over Graphs

From Single-Agent to *Social* Learning

data can be **heterogeneous** across the agents

- $x_{k,t} \in \mathcal{X}_k$

  private streaming data, agent k at time t

- $\ell_k(x_{k,t}|\theta)$

  marginal likelihood, agent k

  (private model)

- $\mu_{k,t} = [\mu_{k,t}(\theta_1), \mu_{k,t}(\theta_2), \ldots, \mu_{k,t}(\theta_M)]$

  belief vector, agent k at time t

Agents can only share beliefs (**not private data**) with their **neighbors**

Joint Bayesian model across the agents **not available**

[ZhaoSayed2012] [JadbabaieMolaviSandroniTahbaz-Salehi2012] [ShahrampourRakhlinJadbabaie2016]

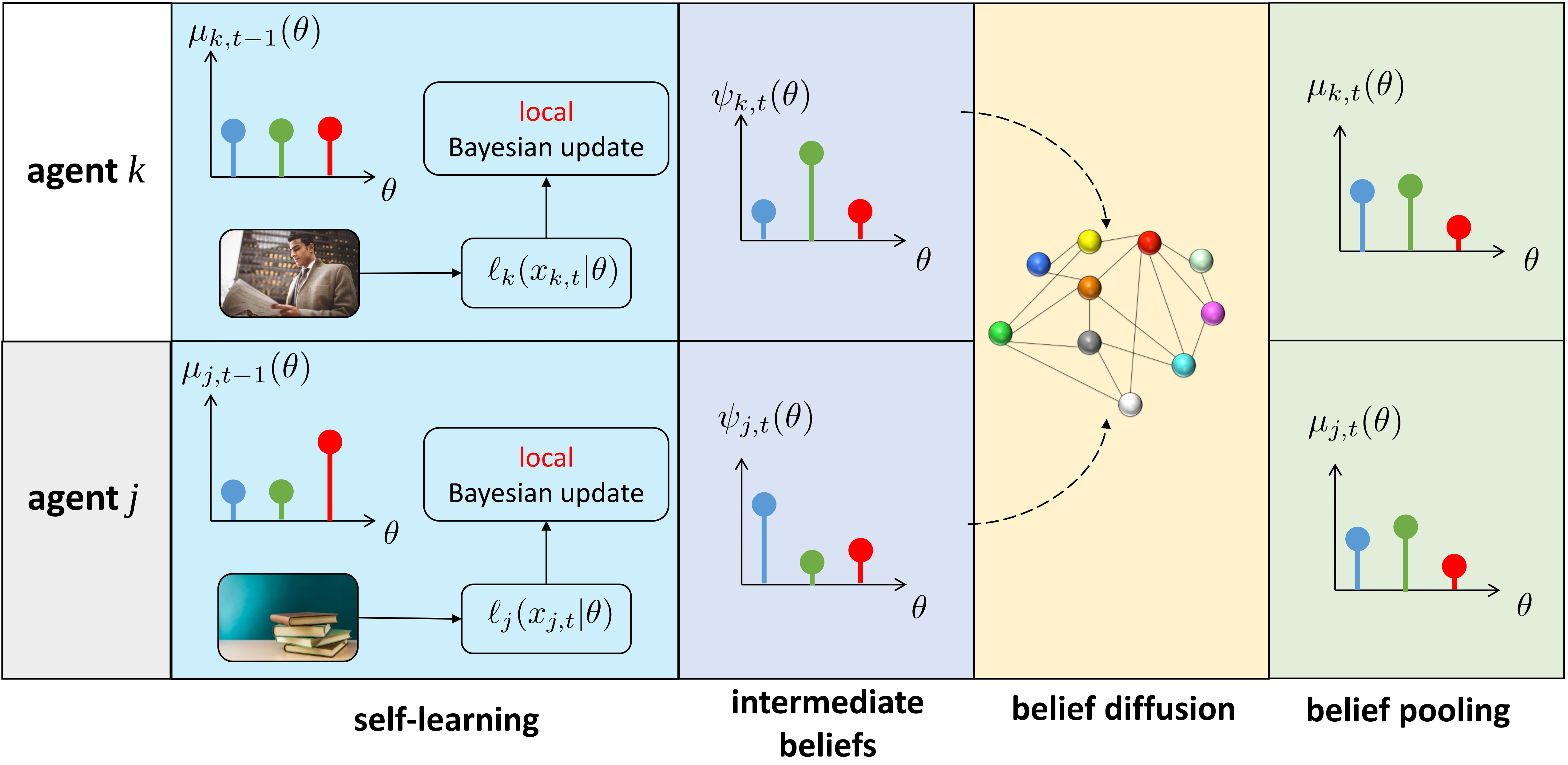
[NedićOlshevskyUribe2017] [MolaviTahbaz-SalehiJadbabaie2018][LalithaJavidiSarwate2018]

Local Bayesian Updates

Each agent builds an **intermediate belief** (to be shared with its neighbors) by updating its previous belief with **its own private likelihood and data**

$$\psi_{k,t}(\theta) \propto \mu_{k,t-1}(\theta)\ell_k(x_{k,t}|\theta)$$

Belief Diffusion Over Graphs



Pooling From Information-Theoretic Viewpoint

[NedićOlshevskyUribe2017] [KolianderEl-LahamDjurićHlawatsch 2022]

- Find a pmf p that **globally** matches the beliefs received from the neighbors

- Minimize a **weighted combination of KL divergences**

global pmf

$$\mu_{k,t} = \arg\min_{p} \sum_{j \in \mathcal{N}_k} a_{jk} D(p || \psi_{j,t})$$

neighbors

Optimal Pooling Rule

$$\mu_{k,t} = \arg\min_{p} \sum_{j \in \mathcal{N}_k} a_{jk} D(p || \psi_{j,t})$$

$$\Downarrow$$

$$\boxed{\mu_{k,t}(\theta) \propto \prod_{j \in \mathcal{N}_k} [\psi_{j,t}(\theta)]^{a_{jk}}}$$

geometric pooling

a.k.a. log-linear pooling

Pooling From Behavioral Viewpoint

[MolaviTahbaz-SalehiJadbabaie2018]

- We can derive the pooling rule from "behavioral" constraints
- Bounded rationality
- Unanimity, monotonicity, independence of irrelevant alternatives,…

The same rule is obtained!!!

Non-Bayesian Social Learning Algorithm

self-learning step

local Bayesian update

$$\psi_{k,t}(\theta) \propto \mu_{k,t-1}(\theta)\ell_k(x_{k,t}|\theta)$$

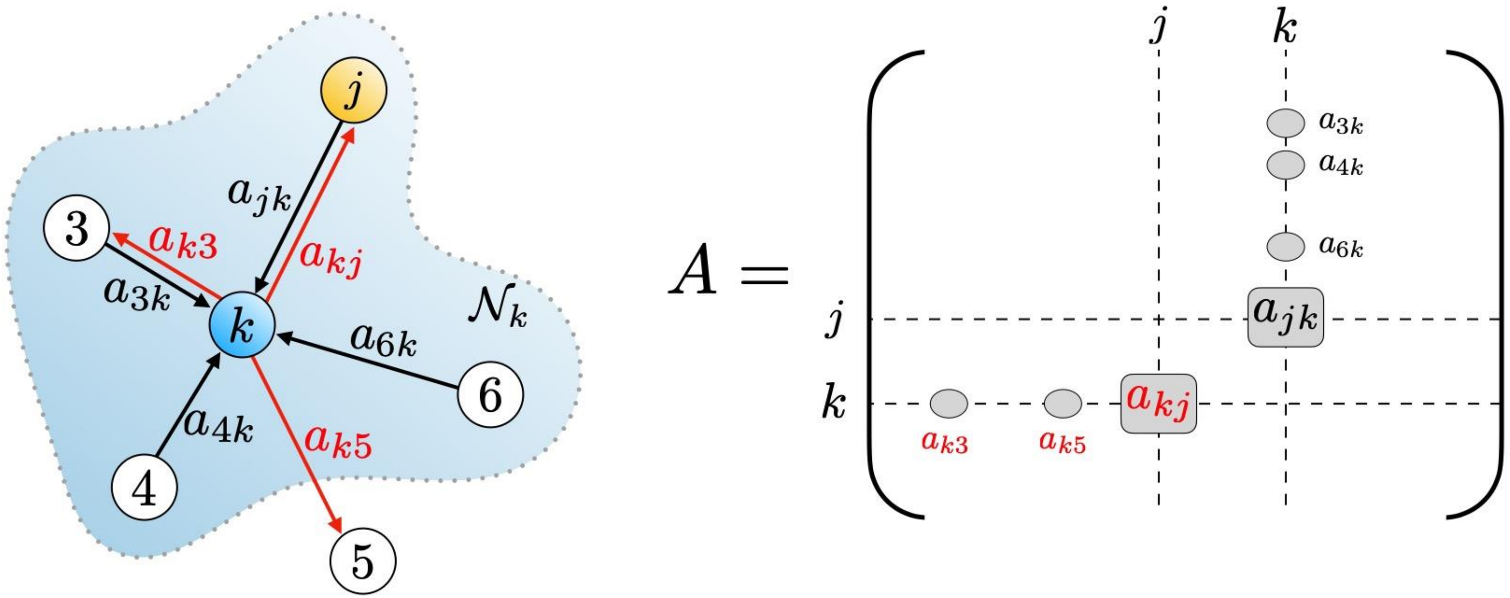
combination step

geometric averaging

$$\mu_{k,t}(\theta) \propto \prod_{j \in \mathcal{N}_k} [\psi_{j,t}(\theta)]^{a_{jk}}$$

Network Graph and Combination Matrix

$$a_{jk} \geq 0, \qquad \sum_{j \in \mathcal{N}_k} a_{jk} = 1$$

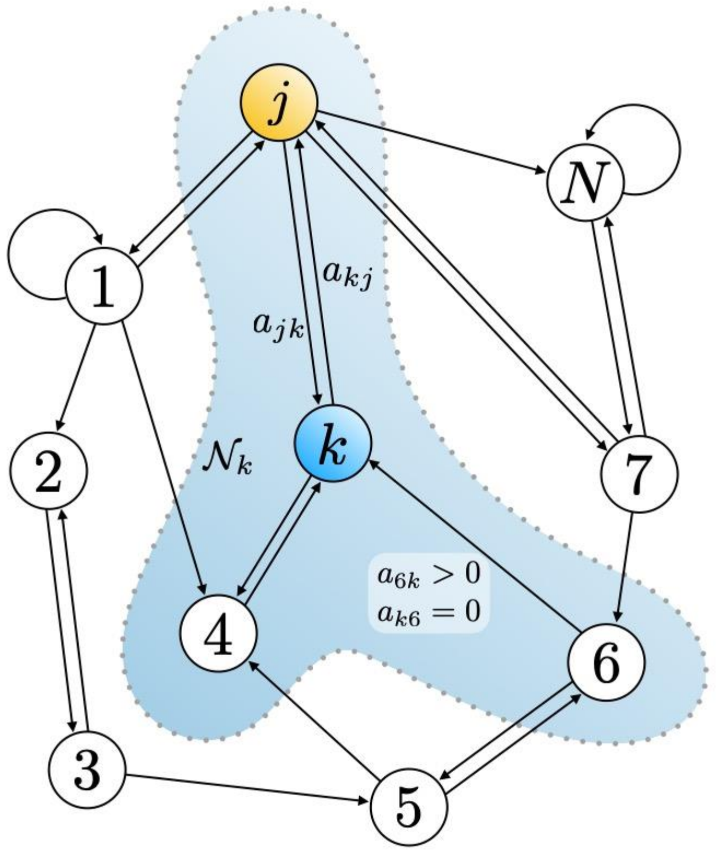


The combination weights and the communication structure involving neighboring agents can be encoded into a **weighted graph**

Strong Graphs

A **path between any two nodes** (in both directions)

Primitive combination matrix



$$\lim_{t\to\infty} A^t = v\mathbb{1}^\top$$

Perron eigenvector

Strong Graphs: Agreement

Perron eigenvector entry

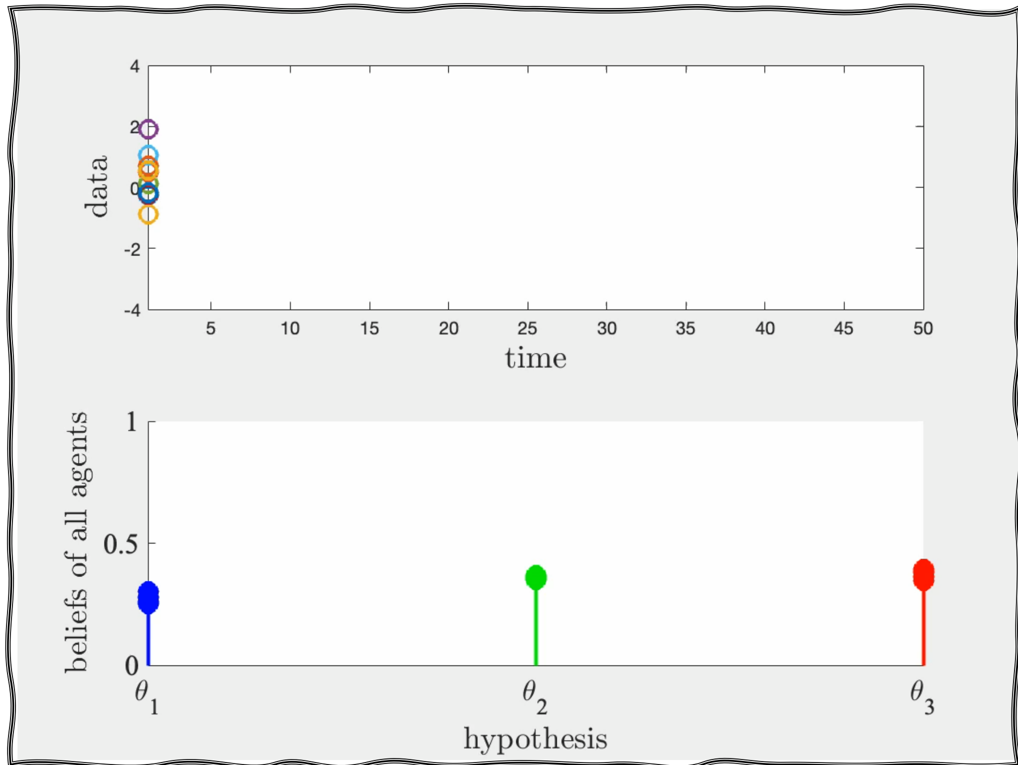
individual divergence replaced by a **network average** divergence

$$D_{\text{net}}(\theta) = \sum_{k=1}^{K} v_k \, D(f_k || \ell_{k,\theta})$$

distribution of agent k

$$\theta^\star = \arg\min_{\theta \in \Theta} D_{\text{net}}(\theta)$$

$$\boldsymbol{\mu}_{k,t}(\theta^\star) \xrightarrow{t\to\infty} 1 \quad \text{almost surely}$$



Strong Graphs: Agreement

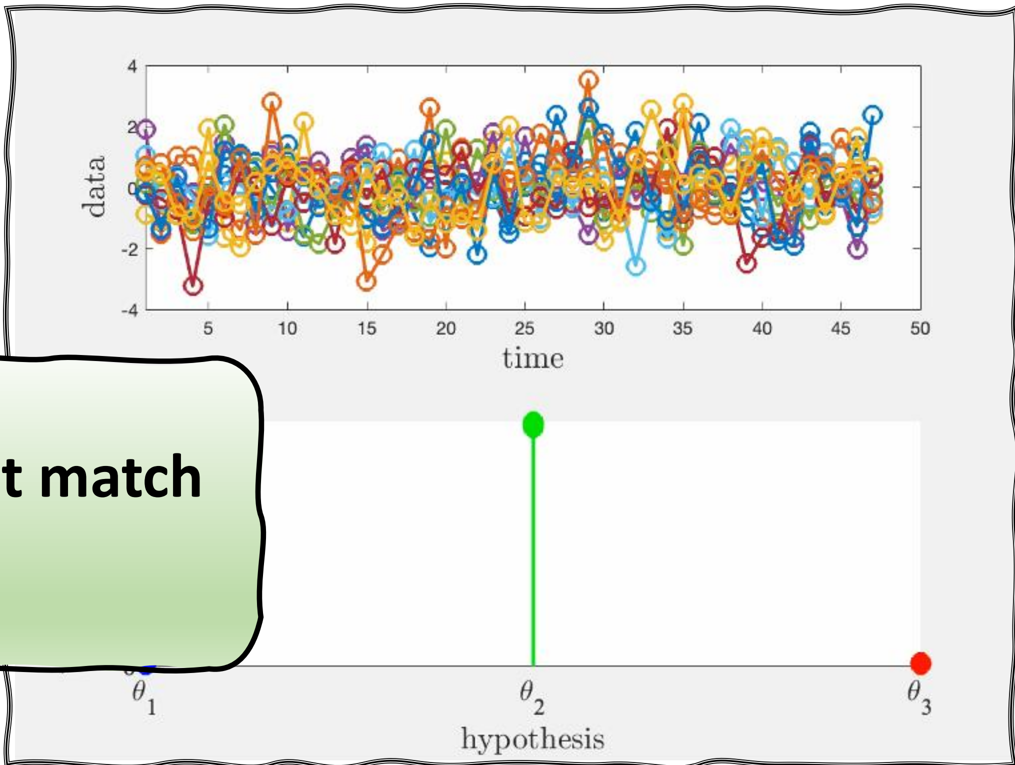
Perron eigenvector entry

$$\theta^\star = \arg\min_{\theta\in\Theta} D_{\rm net}(\theta)$$

$$\boldsymbol{\mu}_{k,t}(\theta^\star) \xrightarrow{t\to\infty} 1 \quad \text{almost surely}$$

individual divergence replaced by a **network average** divergence

$$D_{\rm net}(\theta) = \sum_{k=1}^{K} v_k\, D(f_k || \ell_{k,\theta})$$



The agents agree on the hypothesis providing the best match *on average* across the network

Objective Evidence

Under the objective evidence model, the observations of each agent are drawn from the model ℓ_{k,θ_0} corresponding to a **common true** hypothesis θ_0

$$D_{\text{net}}(\theta) = \sum_{k=1}^{K} v_k \, D(\ell_{k,\theta_0} || \ell_{k,\theta}) > 0$$

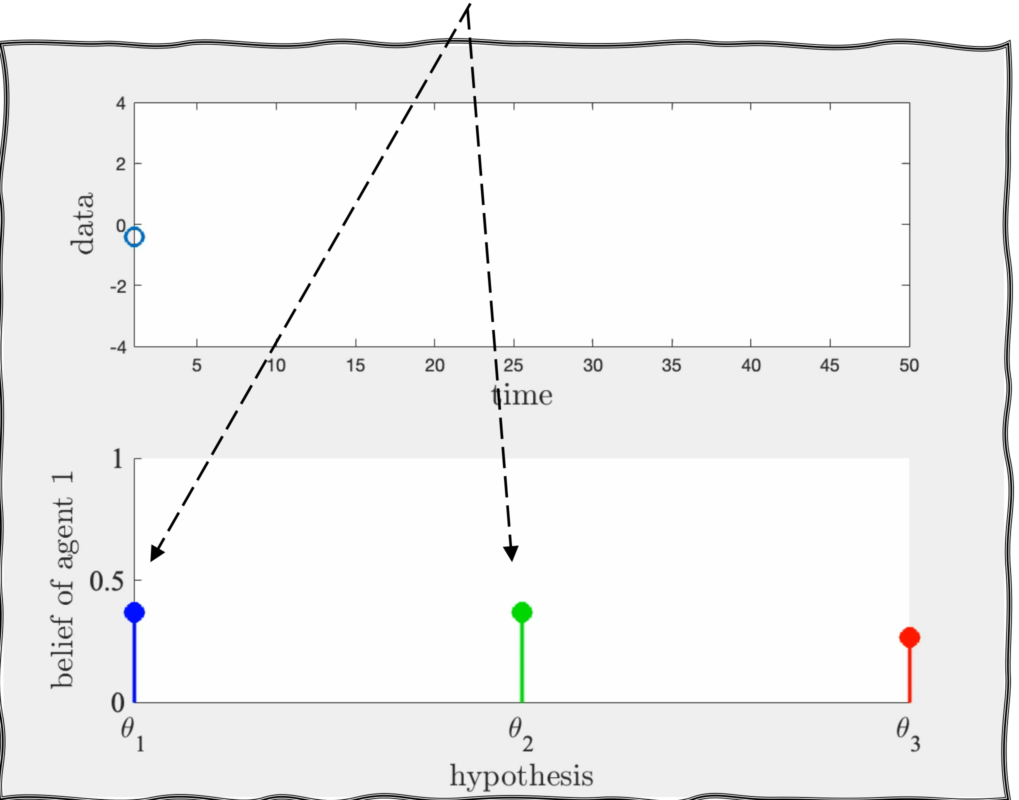
global identifiability

$$\boldsymbol{\mu}_{k,t}(\theta_0) \xrightarrow{t\to\infty} 1 \quad \text{almost surely}$$

Benefits of Cooperation

- The learning accuracy can be improved by combining information from different agents
- Some agents might not be able to solve the problem on their own (**lack of local identifiability**)

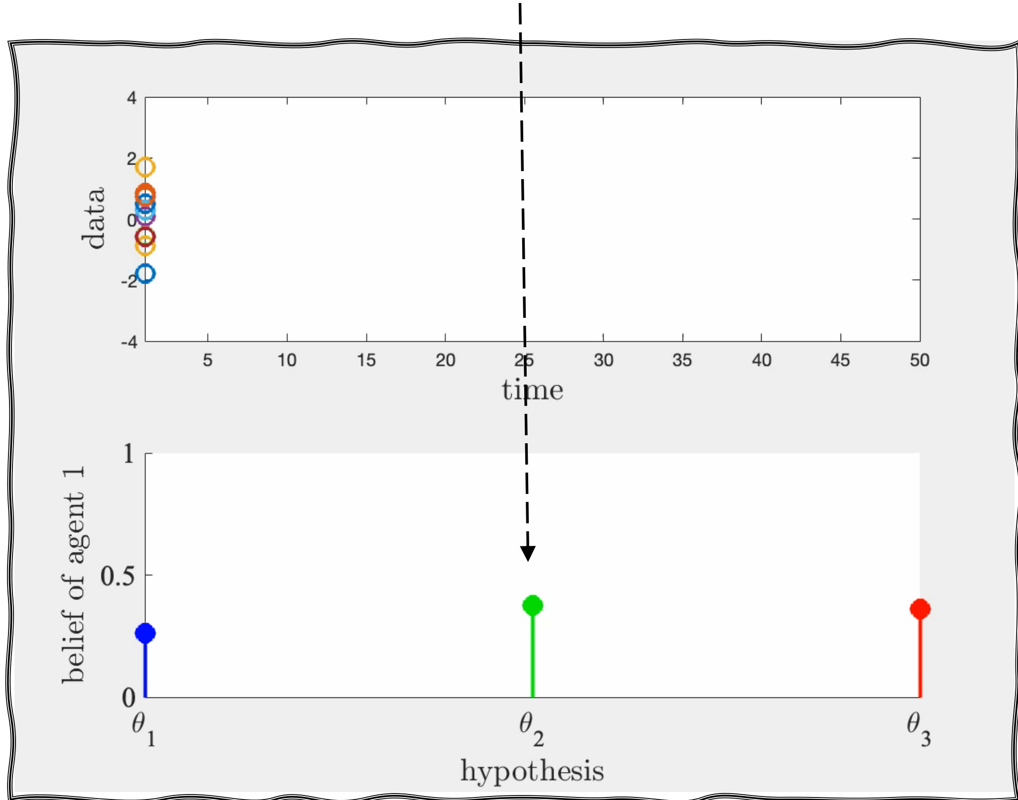
Agent 1 cannot distinguish θ_1 from θ_2



true hypothesis

θ_2

Agent 1 learns well through social learning



Subjective Evidence and Fake News

Under the subjective evidence model, different agents can have different underlying hypotheses

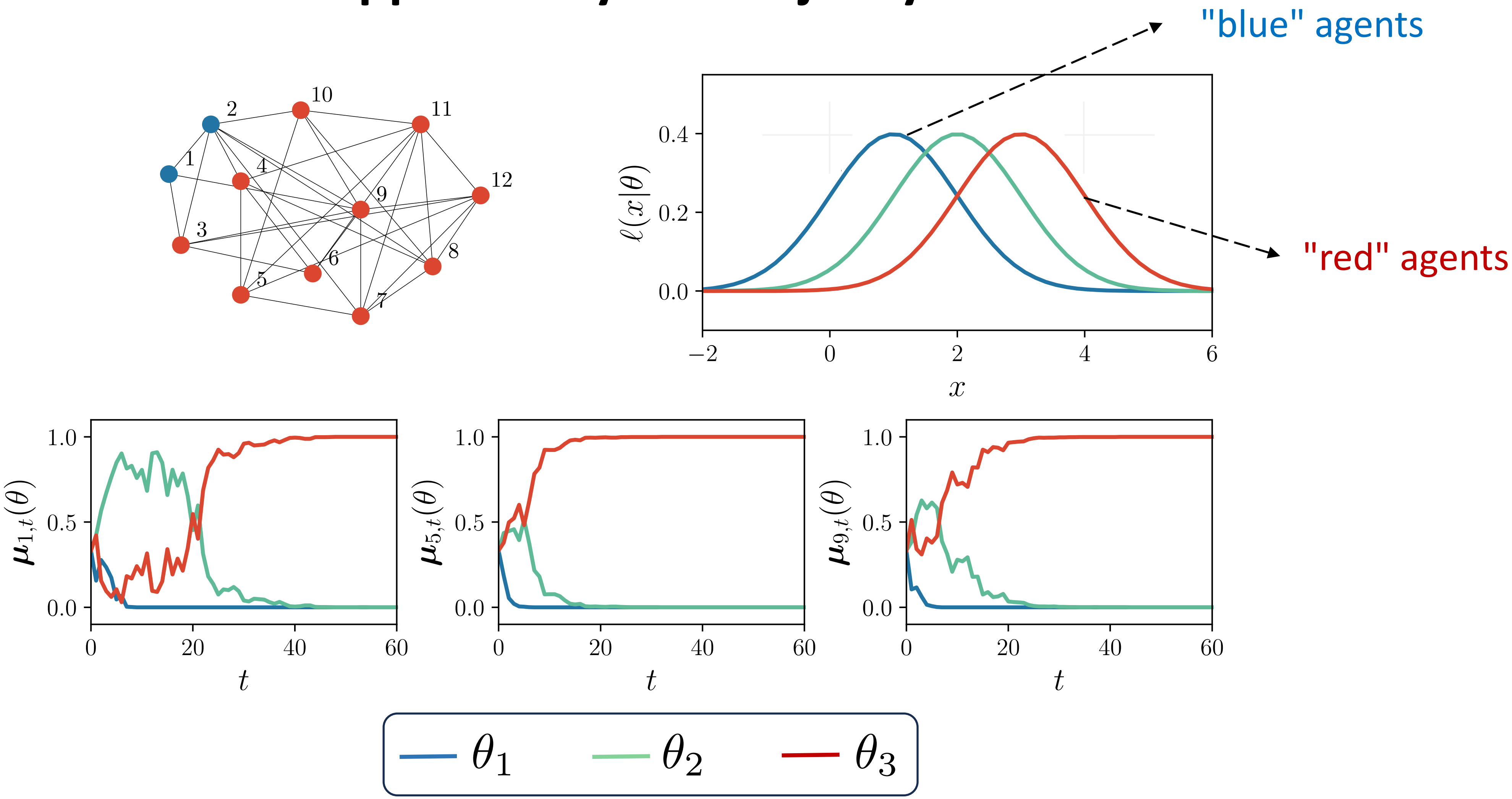
$$D_{\text{net}}(\theta) = \sum_{k=1}^{K} v_k \, D(\ell_{k,\theta_k} || \ell_{k,\theta})$$

$$\theta^\star = \arg\min_{\theta \in \Theta} D_{\text{net}}(\theta)$$

But in this case…the agents agree on which hypothesis?

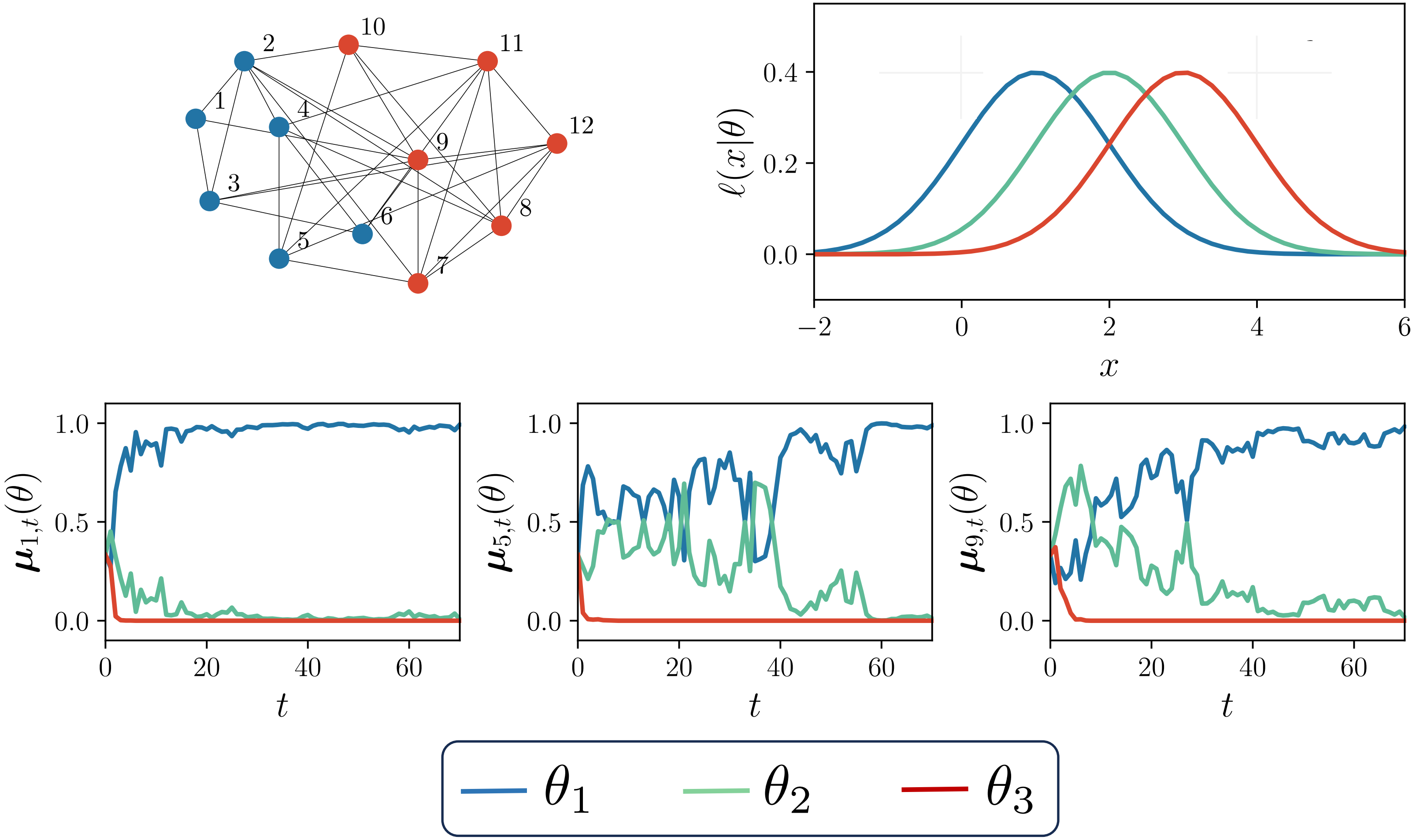
Majority Builds a Common Opinion

Here all agents place full mass on the hypothesis
supported by the majority



Centrality Builds a Common Opinion

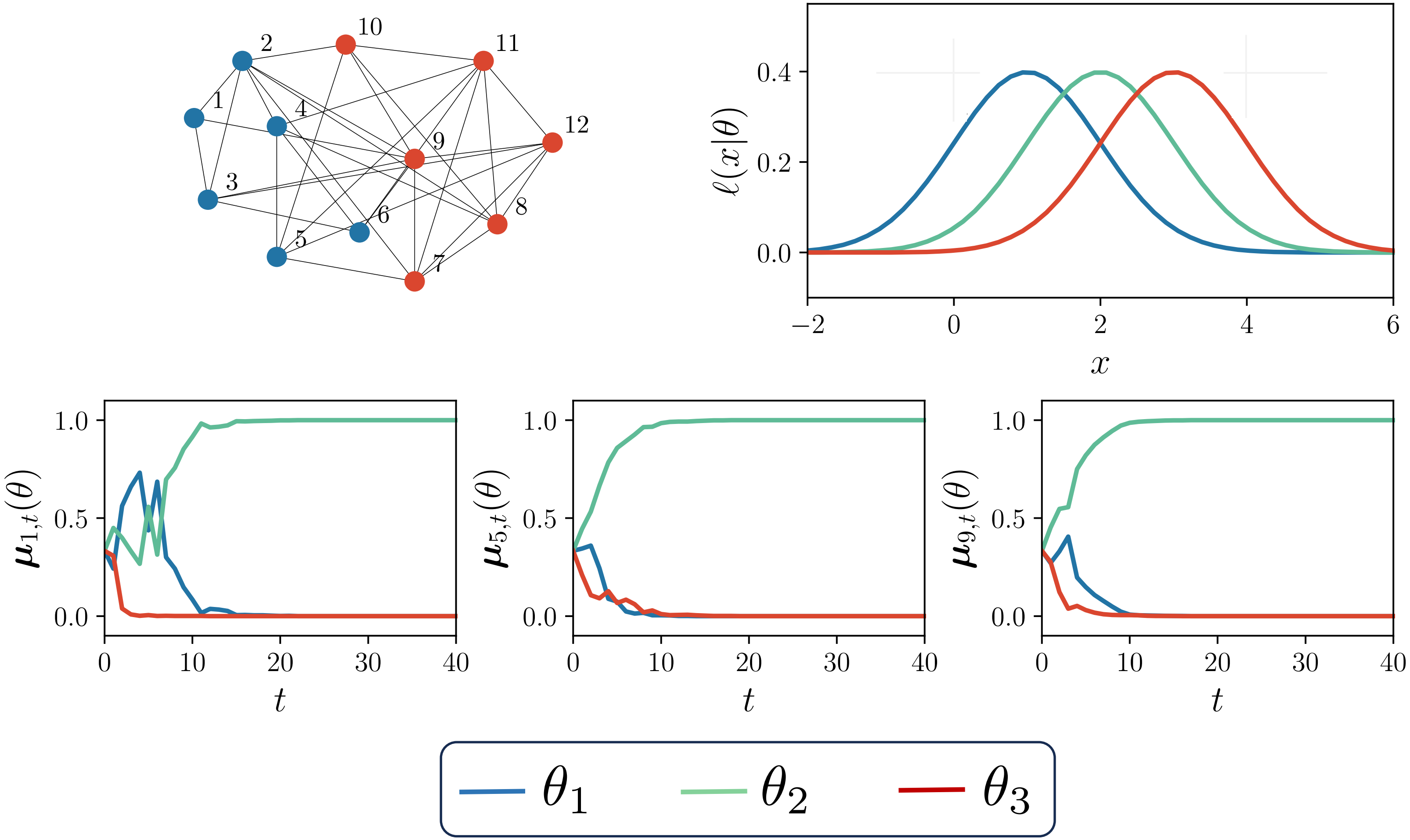
Here all agents place full mass on the hypothesis
supported by the agents with more neighbors



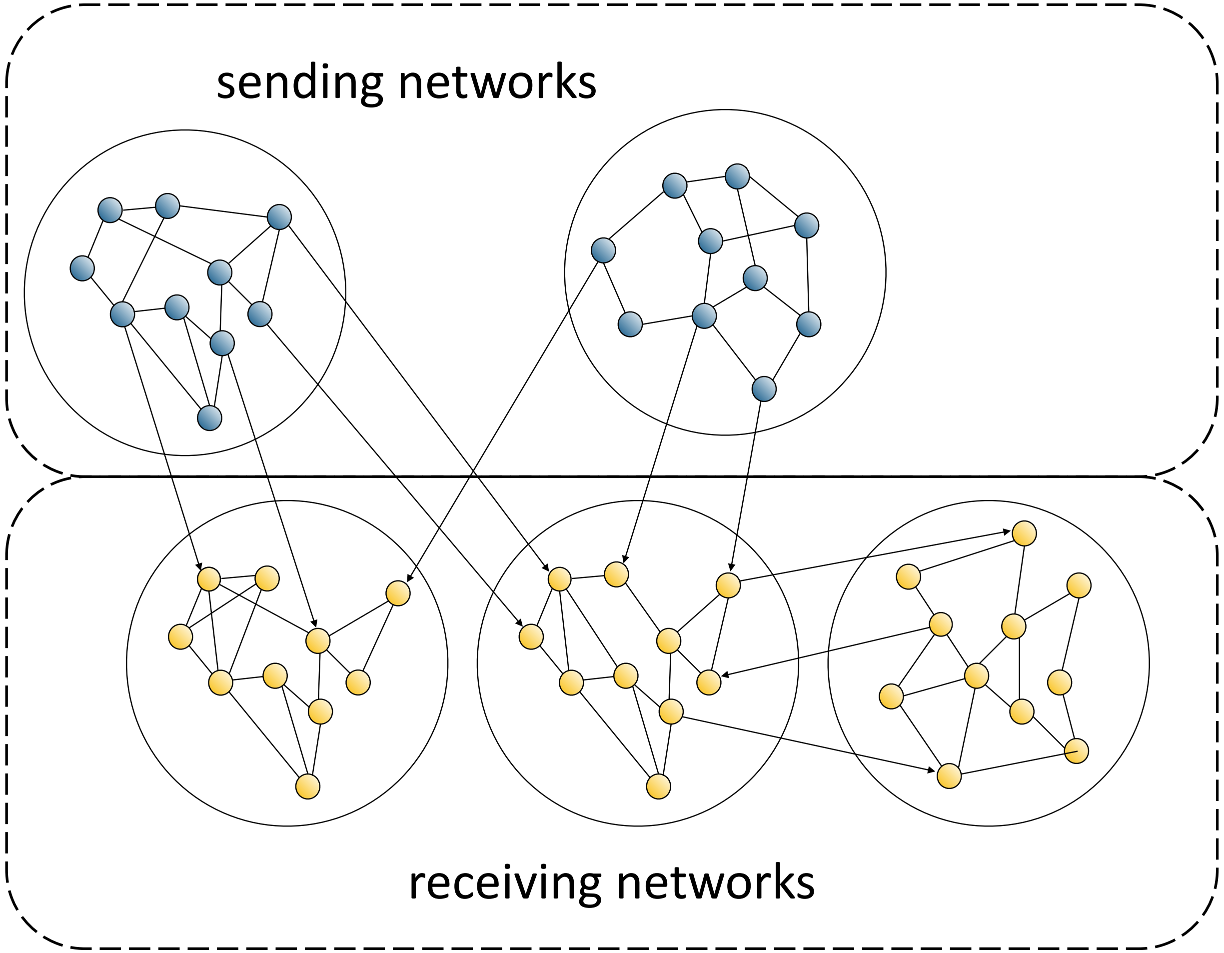
Truth is Somewhere in Between

Here half network says θ_1, the other half says θ_3

All agents opt for θ_2!



Weak Graphs



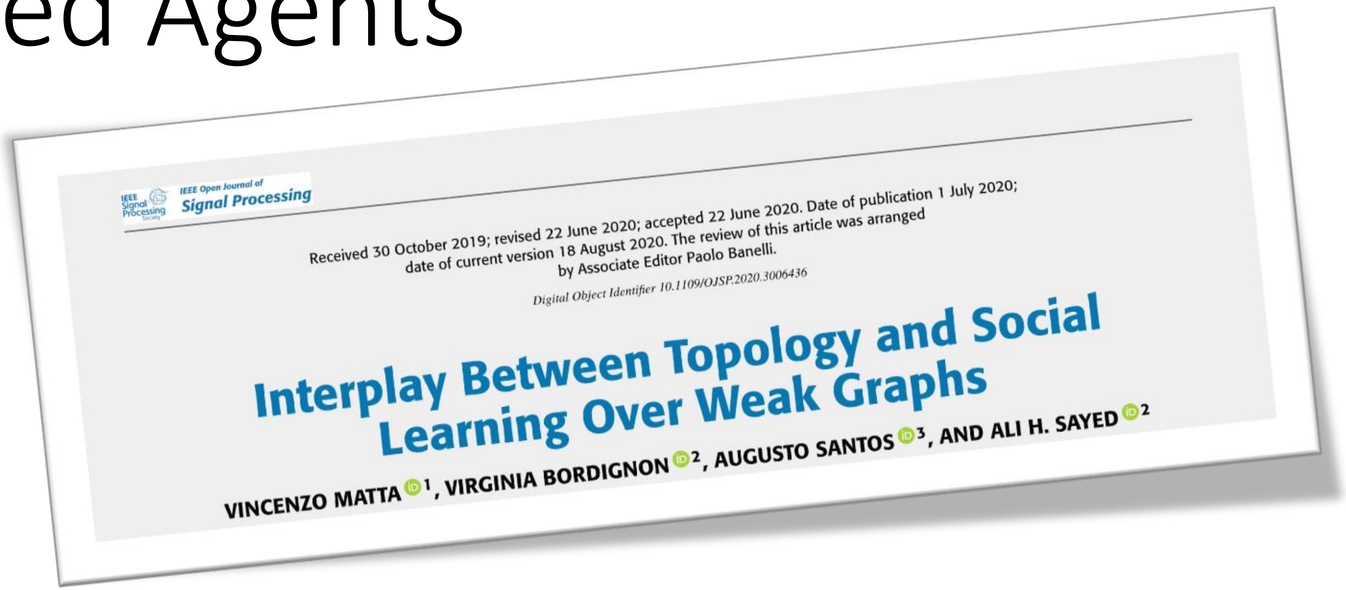
- **Any** graph that is not strong can be represented in a **canonical** form where it is partitioned into **sending** and **receiving** networks

- Useful real-world examples:
  - celebrities over social networks
  - media networks

$$A = \left[\begin{array}{c|c} A_{\mathcal{S}} & A_{\mathcal{S}\mathcal{R}} \\ \hline 0 & A_{\mathcal{R}} \end{array}\right]$$

Influencers vs. Influenced Agents

$$\lim_{t\to\infty} A^t = \left[\begin{array}{c|c} V & W \\ \hline 0 & 0 \end{array}\right]$$


IEEE Open Journal of Signal Processing

Received 30 October 2019; revised 22 June 2020; accepted 22 June 2020. Date of publication 1 July 2020; date of current version 18 August 2020. The review of this article was arranged by Associate Editor Paolo Banelli.

Digital Object Identifier 10.1109/OJSP.2020.3006436

Interplay Between Topology and Social Learning Over Weak Graphs

VINCENZO MATTA[1], VIRGINIA BORDIGNON[2], AUGUSTO SANTOS[3], AND ALI H. SAYED[2]

$$D_k(\theta) = \sum_{j\in\mathcal{S}} w_{jk} D(f_j||\ell_{j,\theta})$$

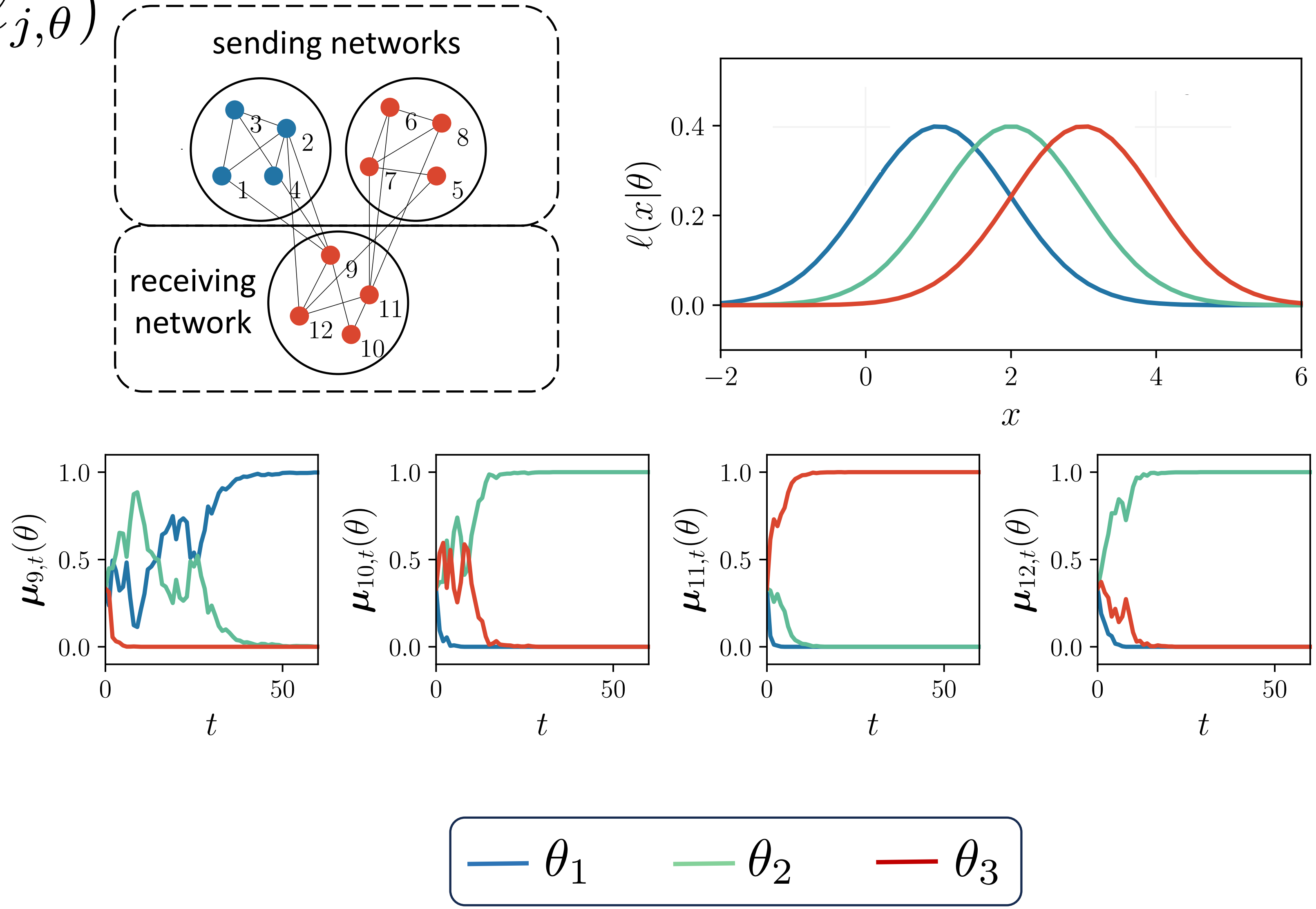
only sending networks!!!

The sending networks exert a domineering role (**influencers**) over the receiving networks (**influenced**)

Weak Graphs: Discord

$$D_{k}(\theta) = \sum_{j \in \mathcal{S}} w_{jk}\, D(f_j || \ell_{j,\theta})$$

agent dependent!!!



Part III

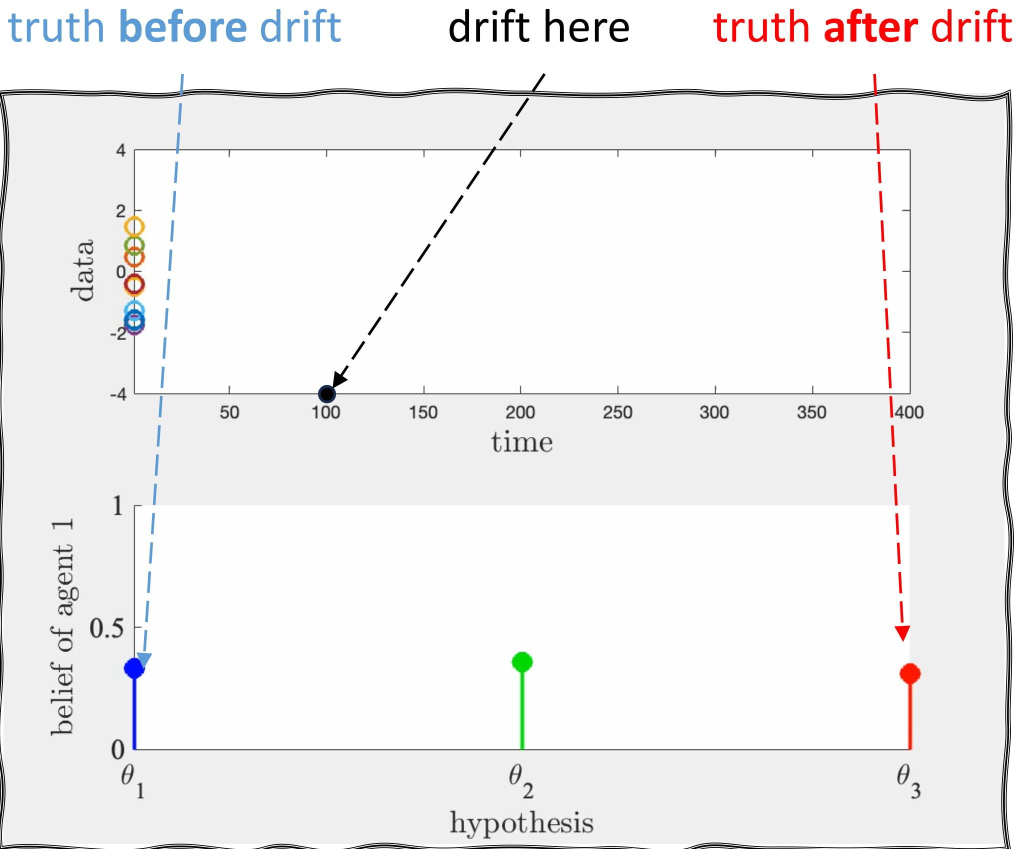
Recent Trends in Social Learning

Adaptive Social Learning

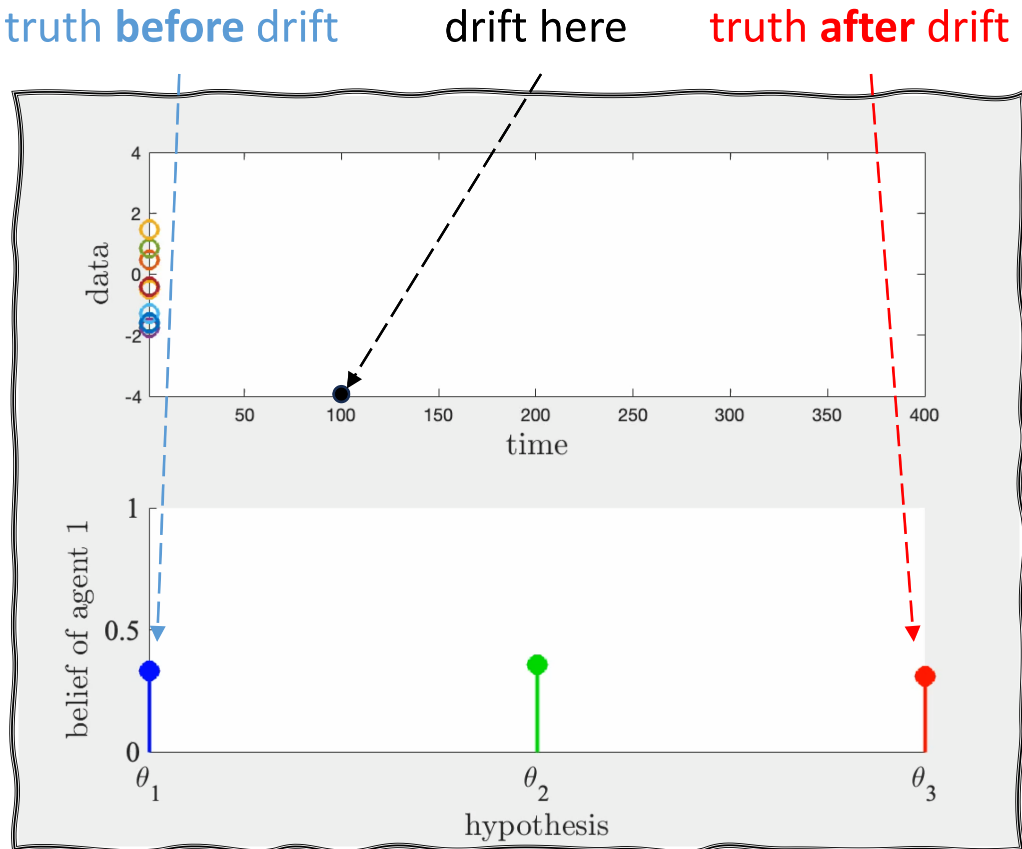
Stubbornness vs. Adaptation

Fluctuations keep adaptation alive

Traditional Social Learning



Adaptive Social Learning



Adaptive Social Learning


IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 67, NO. 9, SEPTEMBER 2021

6053

Adaptive Social Learning

Virginia Bordignon, *Graduate Student Member, IEEE*, Vincenzo Matta, *Member, IEEE*, and Ali H. Sayed, *Fellow, IEEE*

traditional
Bayesian update

$$\psi_{k,t}(\theta) = \arg\min_{p} \left\{ (1-\delta) D(p || \mu_{k,t}^{\text{Bayes}}) + \delta D(p || \mu_{k,t}^{\text{no-past}}) \right\}$$

adaptation
parameter

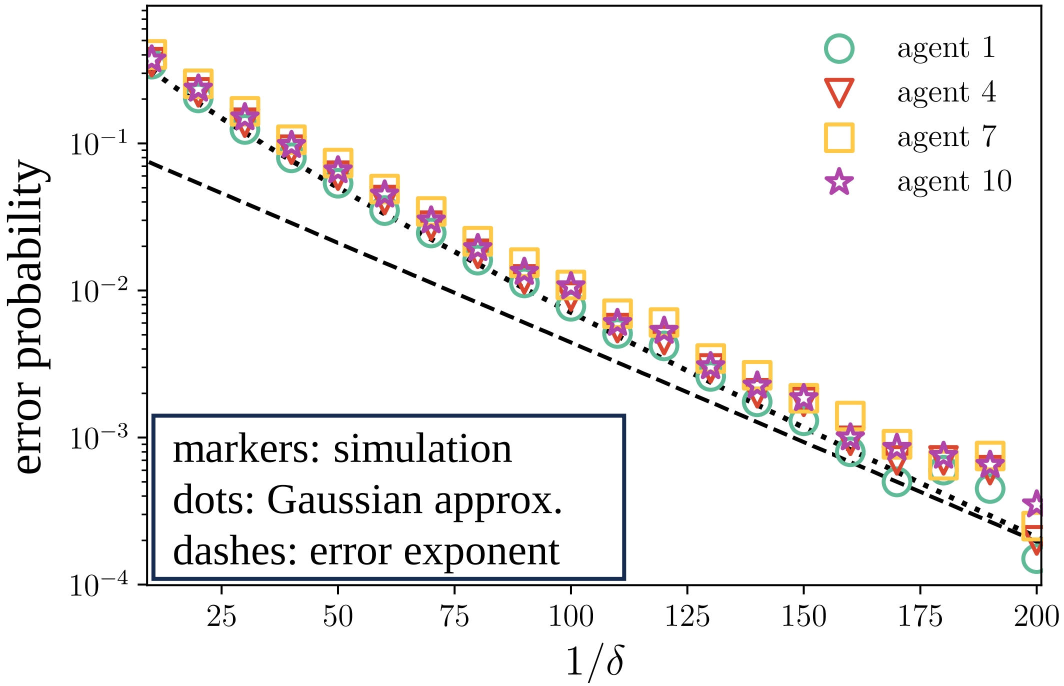
Bayesian update **ignoring the past belief**

$$\psi_{k,t}(\theta) \propto \mu_{k,t-1}^{1-\delta}(\theta) \ell_k(x_{k,t} | \theta)$$

Adaptation and Learning

Fundamental adaptation/learning trade-off

$$\text{Prob[error]} \sim \exp\{-\text{adaptation time}\}$$



(prop. to adaptation time)

theoretical analysis: weak law of slow adaptation, asymptotic normality, large deviations,…

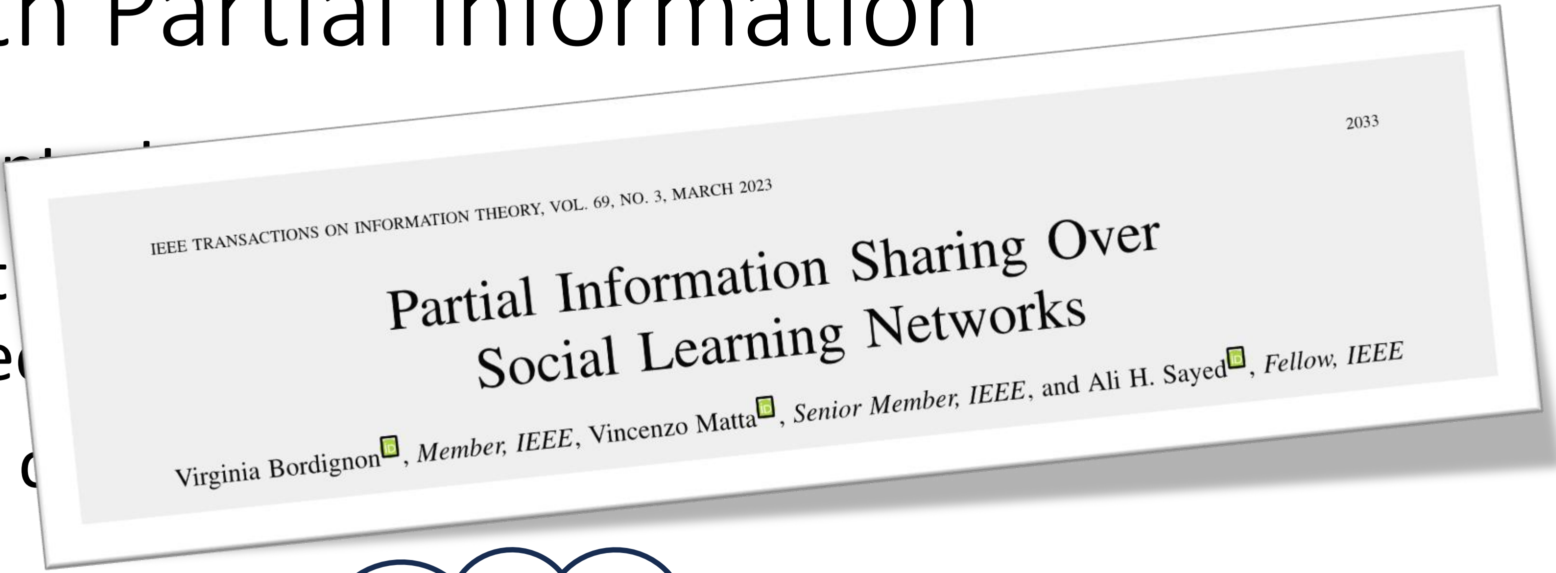
Social Learning With Partial Information

Social Learning With Partial Information

- In many applications, the agents share **partial opinions**
- For example, the agents want to form their opinions on some product brands, but they talk only about a specific one
- How the learning mechanism changes?

Social Learning With Partial Information

- In many applications, the age…
- For example, the agents want … but they talk only about a spe…
- How the learning mechanism …

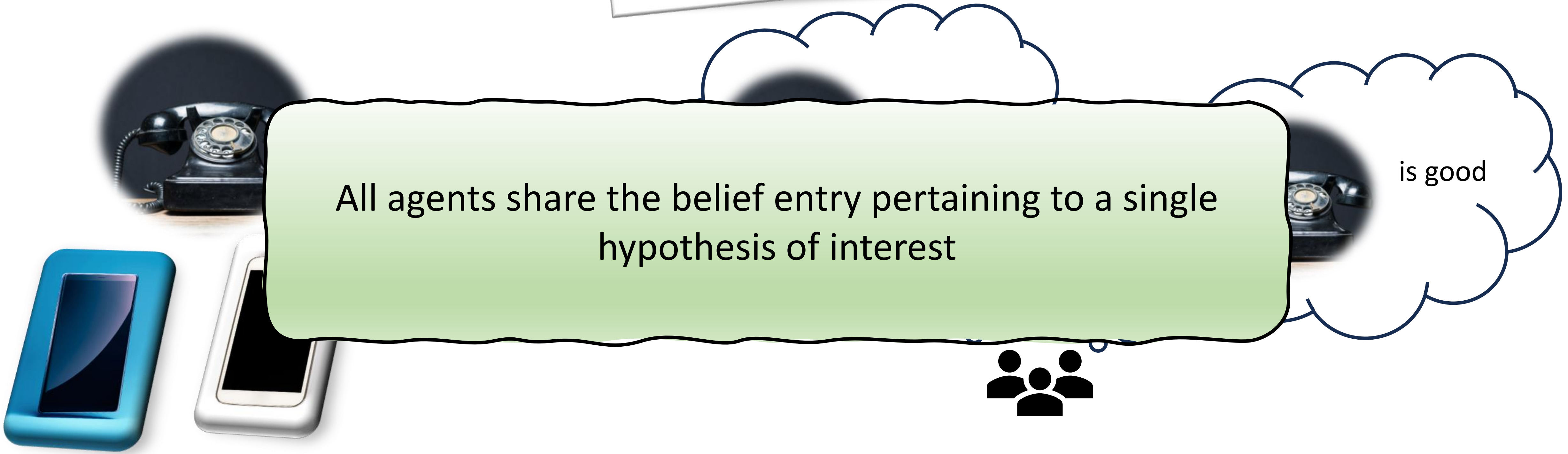


IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 69, NO. 3, MARCH 2023

2033

Partial Information Sharing Over Social Learning Networks

Virginia Bordignon, *Member, IEEE*, Vincenzo Matta, *Senior Member, IEEE*, and Ali H. Sayed, *Fellow, IEEE*



All agents share the belief entry pertaining to a single hypothesis of interest

Filling Strategy

hypothesis of interest

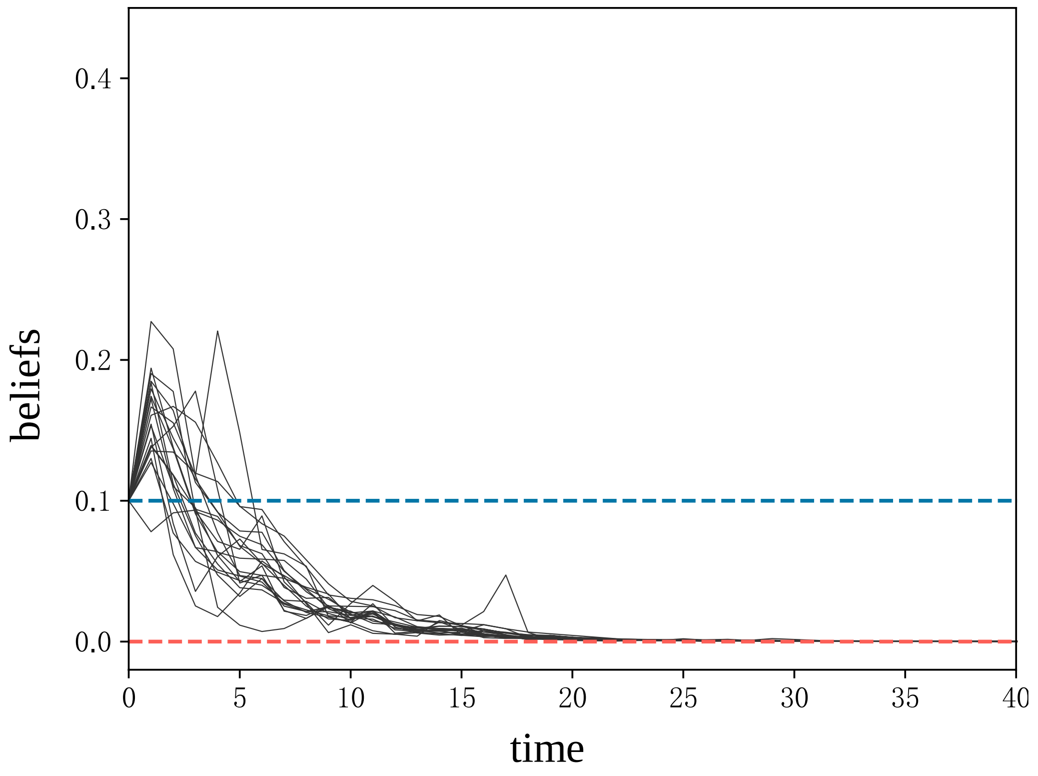
- Agent k receives from its neighbors only the belief pertaining to $\theta^\bullet$
- Fill in the belief entries for the complementary set $\mathcal{T} \triangleq \left\{\theta \in \Theta : \theta \neq \theta^\bullet\right\}$
- **Bayesian filling** strategy $\widehat{\psi}_{j,t}^{(k)}(\theta) = p_k(\theta|\mathcal{T})\left[1 - \psi_{j,t}(\theta^\bullet)\right]$
- agent k uses its **most updated knowledge** stored in its belief $\psi_{k,t}(\theta)$

conditional belief given that the hypothesis is not $\theta^\bullet$

$$p_k(\theta|\mathcal{T}) = \frac{\psi_{k,t}(\theta)}{1 - \psi_{k,t}(\theta^\bullet)}$$

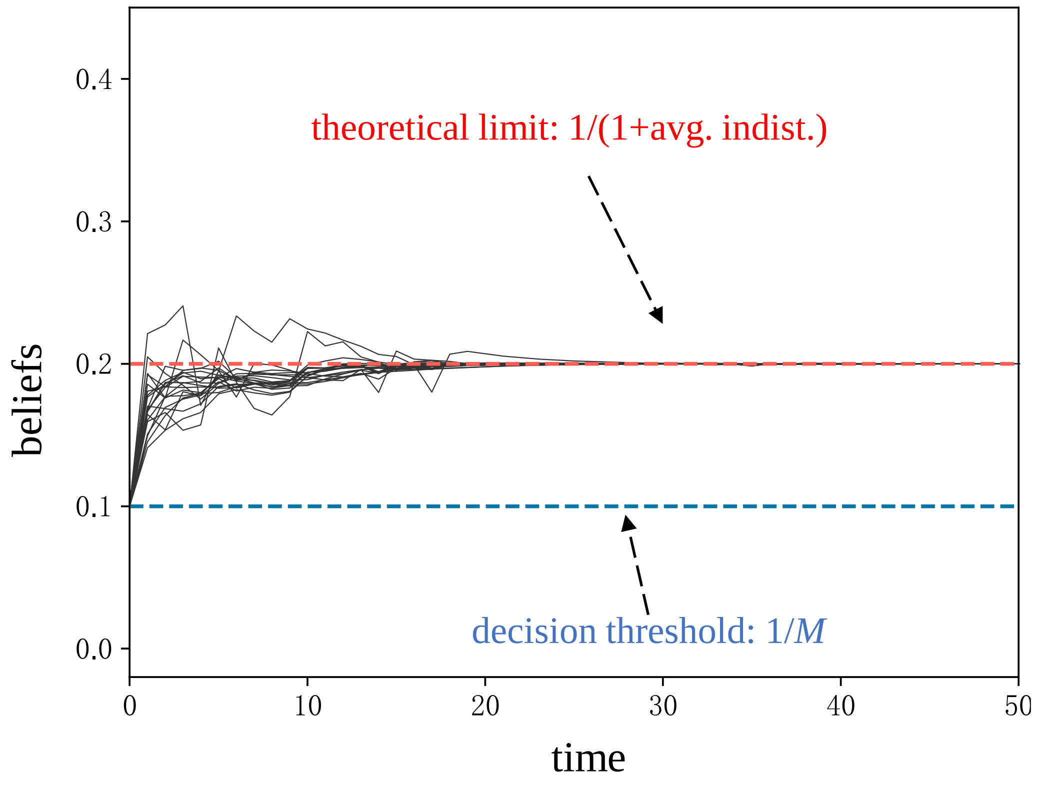
Learning With Partial Information

The hypothesis of interest is false



The hypothesis of interest is **correctly rejected**

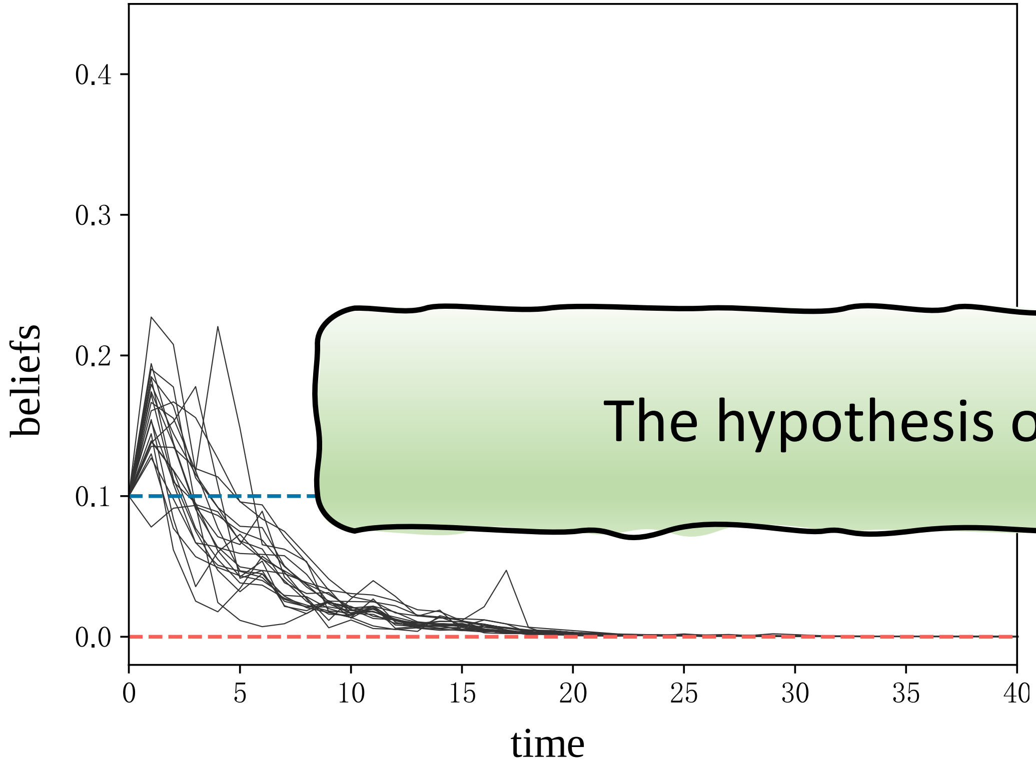
The hypothesis of interest is true



The beliefs of the true hypothesis converge to a positive value. There exists a **decision threshold that implies truth learning**

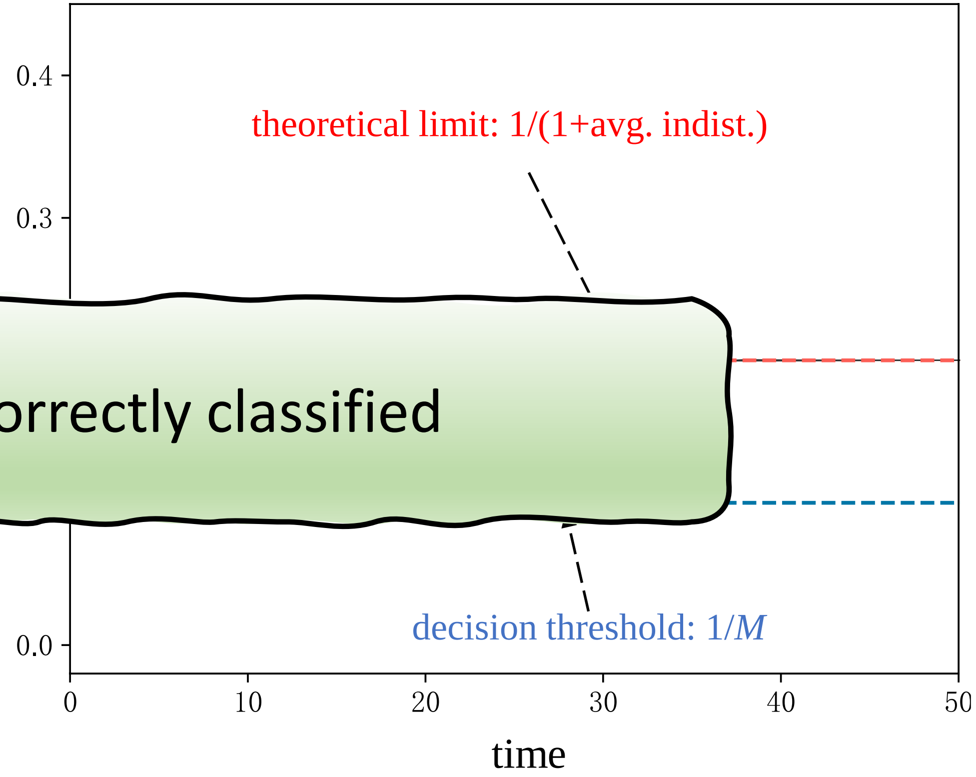
Learning With Partial Information

The hypothesis of interest is false



The hypothesis of interest is **correctly rejected**

The hypothesis of interest is true



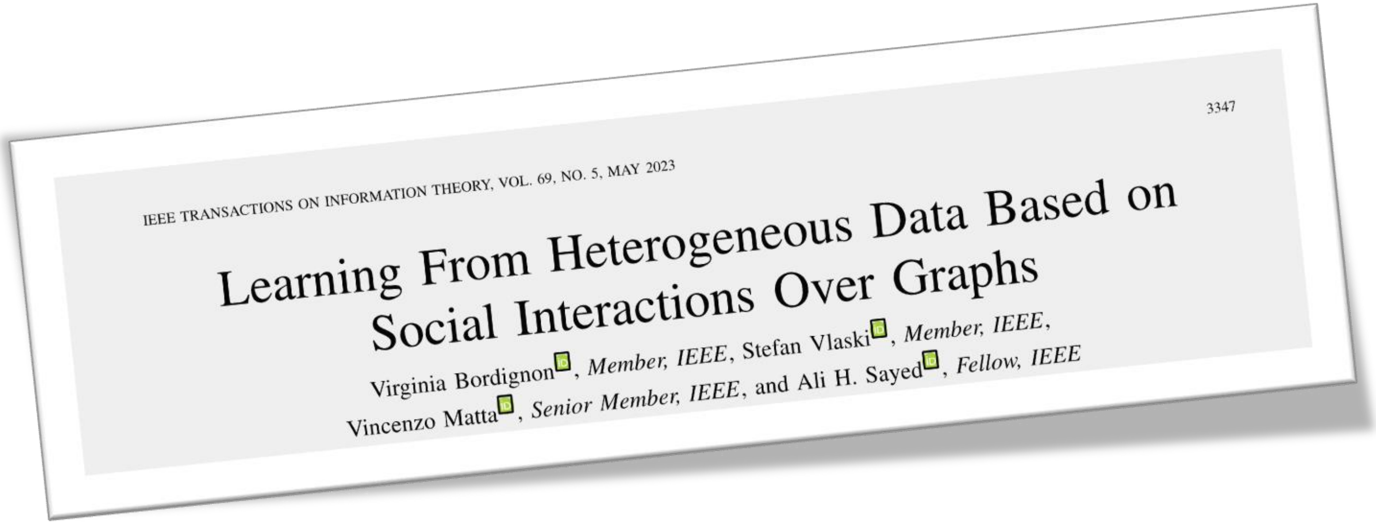
The beliefs of the true hypothesis converge to a positive value. There exists a **decision threshold that implies truth learning**

The hypothesis of interest is correctly classified

Social Machine Learning

Social Machine Learning

- **Outer layer: Training phase**
  Each agent builds its own models from some clues
- **Inner layer: Prediction phase**
  All agents run social learning algorithms with the learned models

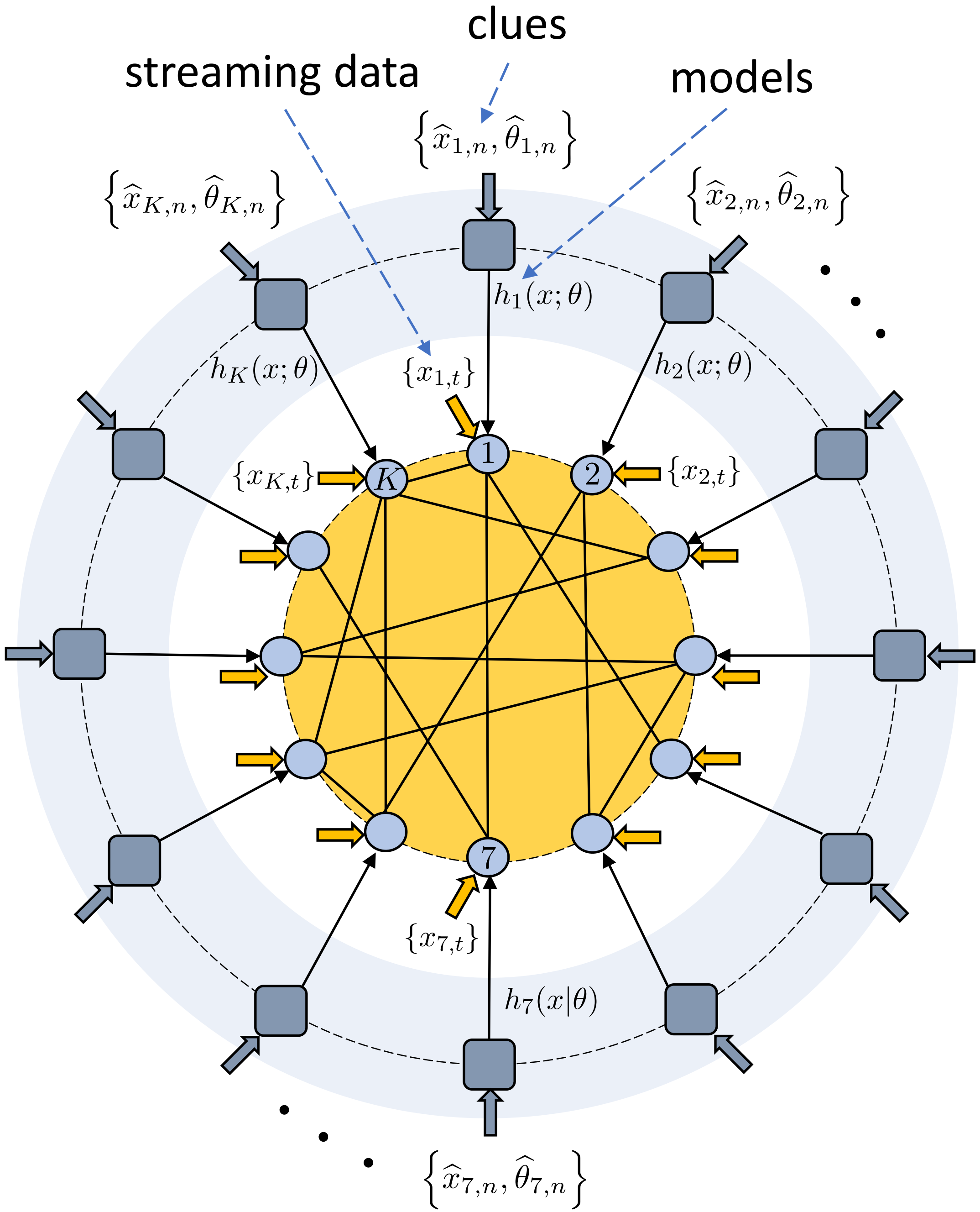


3347

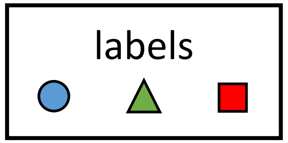
IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 69, NO. 5, MAY 2023

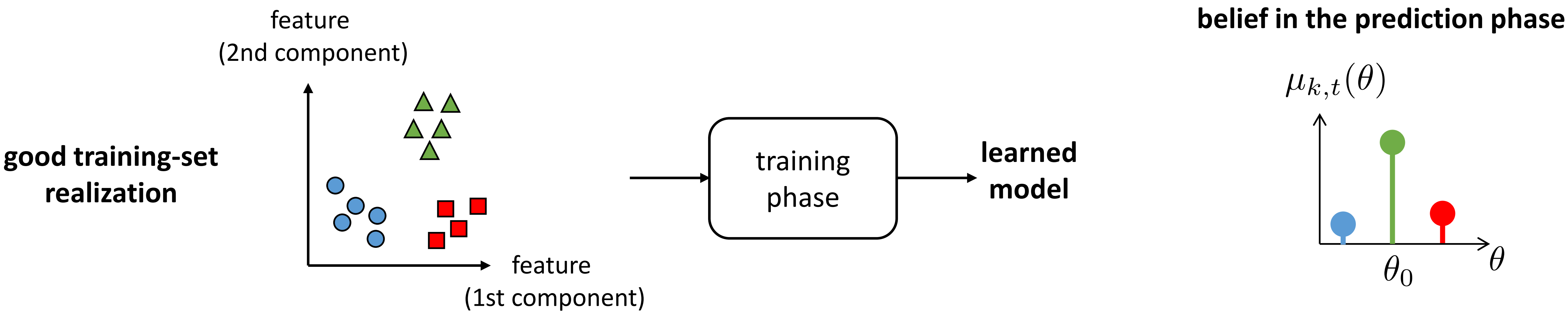
Learning From Heterogeneous Data Based on Social Interactions Over Graphs

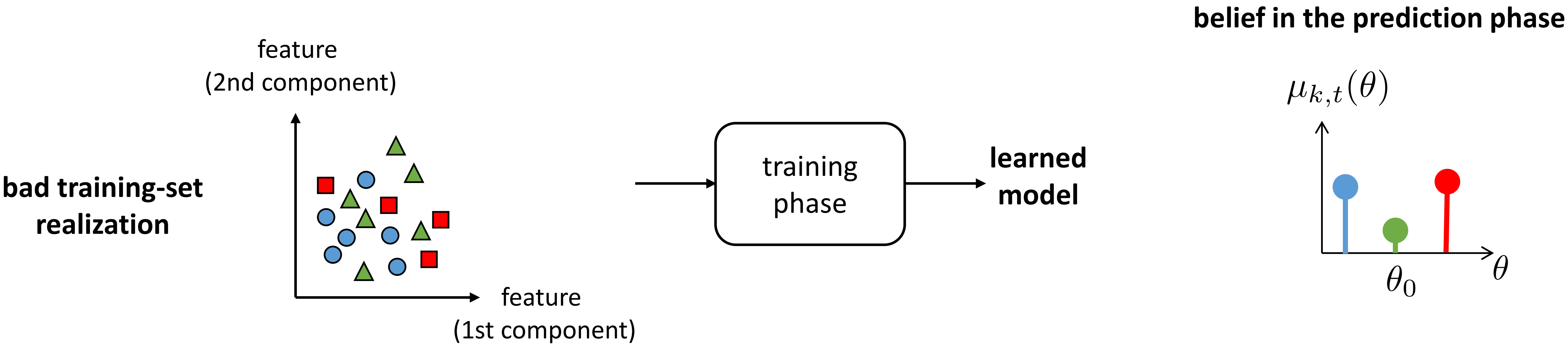
Virginia Bordignon, *Member, IEEE*, Stefan Vlaski, *Member, IEEE*, Vincenzo Matta, *Senior Member, IEEE*, and Ali H. Sayed, *Fellow, IEEE*



Training sets







Consistent Learning

What is the probability that the training set yields good decision models?

$$\mathscr{E}(R_{\mathrm{net}}) \approx 0.2812\left(1 - \frac{R_{\mathrm{net}}}{\log 2}\right)$$

$$\mathrm{Prob}[\text{consistent learning}] \geq 1 - 2\exp\left\{-\frac{N_{\mathrm{clues}}}{\gamma^2}\left(\mathscr{E}(R_{\mathrm{net}}) - \rho_{\mathrm{net}}\right)^2\right\}$$

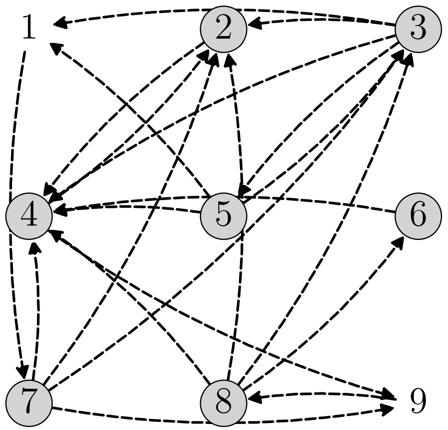
max. no. of training examples

parameter accounting for discrepancies across the agents

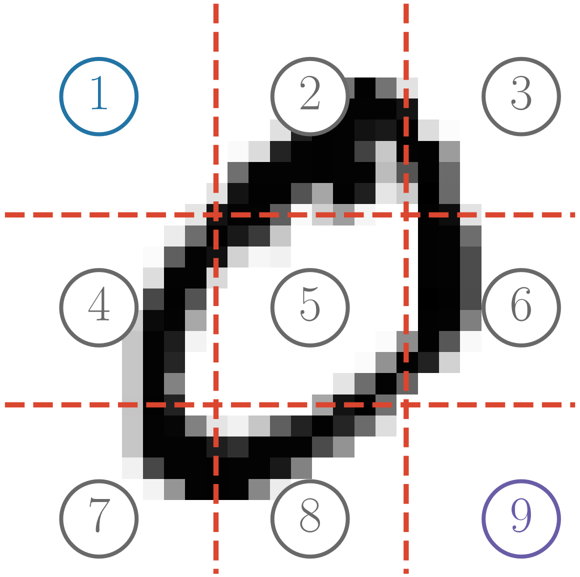
global risk

Rademacher complexity

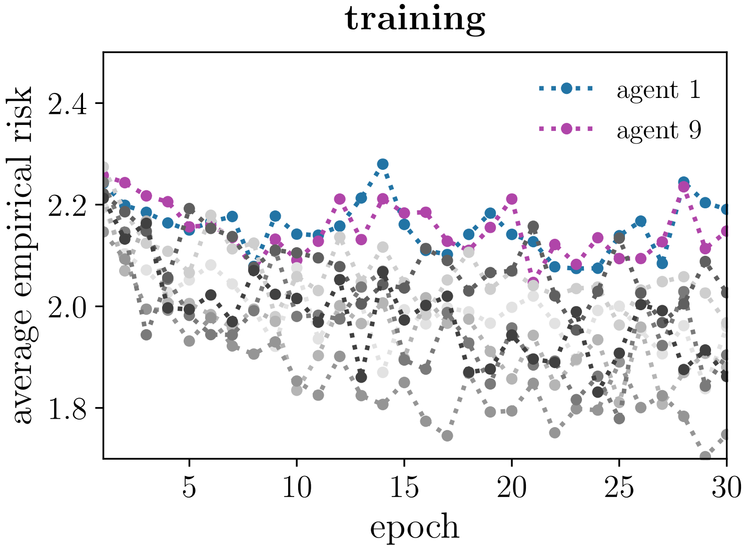
Social Machine Learning Example

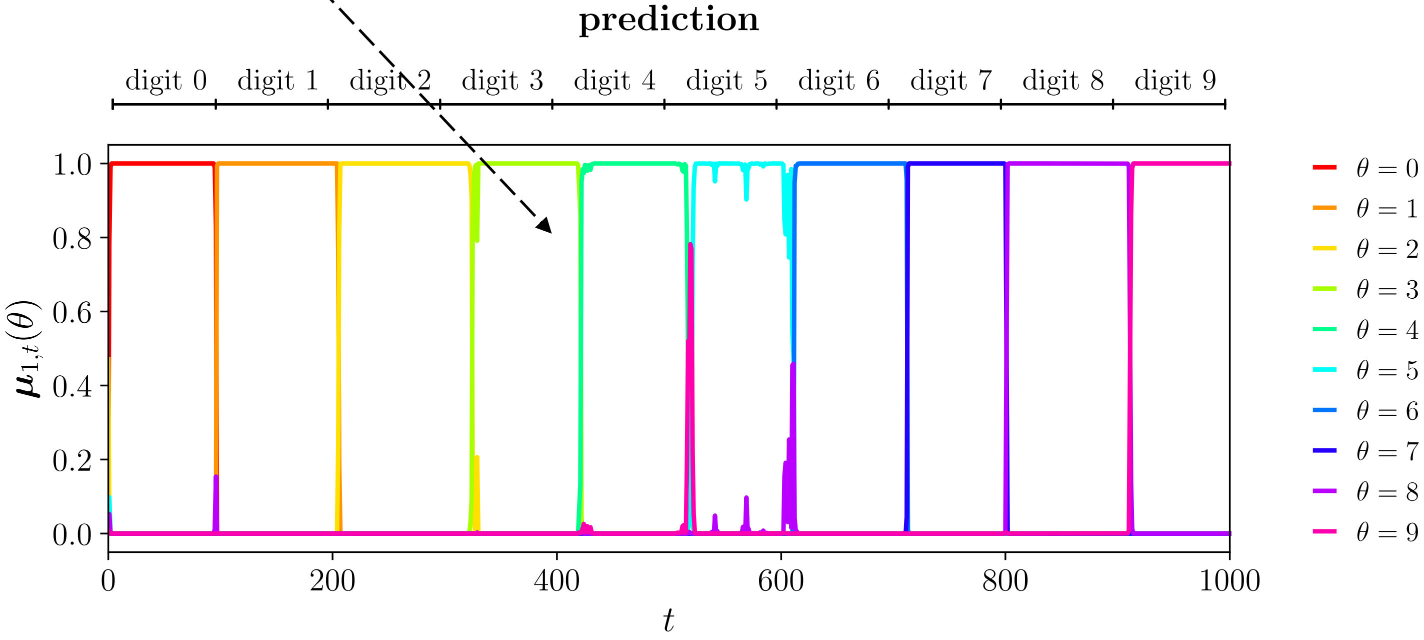


different agents observe different portions of a "digit" image



Digits are correctly predicted with social machine learning





References

- R. H. Berk. *Limiting behavior of posterior distributions when the model is incorrect*. The Annals of Mathematical Statistics, 1966.
- K. Friston, J. Kilner, and L. Harrison. *A free energy principle for the brain*. Journal of physiology-Paris, 2006.
- X. Zhao and A. H. Sayed. *Learning over social networks via diffusion adaptation.* Proc. Asilomar 2012.
- A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi. *Non-Bayesian social learning*. Games and Economic Behavior, 76(1), 2012.
- S. Shahrampour, A. Rakhlin and A. Jadbabaie. *Distributed Detection: Finite-Time analysis and impact of network topology.* IEEE Trans. Automatic Control, 2016.
- A. Nedić, A. Olshevsky, and C. A. Uribe. *Fast convergence rates for distributed non-Bayesian learning*. IEEE Trans. Automatic Control, 2017.
- P. Molavi, A. Tahbaz-Salehi, and A. Jadbabaie. *A theory of non-Bayesian social learning*. Econometrica, 2018.
- A. Lalitha, T. Javidi, and A. D. Sarwate. *Social learning and distributed hypothesis testing*. IEEE Trans. Information Theory, 2018.
- V. Matta, V. Bordignon, A. Santos and A. H. Sayed. *Interplay between topology and social learning over weak graphs.* IEEE Open Journal of Signal Processing, 2020.
- V. Bordignon, V. Matta and A. H. Sayed. *Adaptive social learning.* IEEE Trans. Information Theory, 2021.
- G. Koliander, Y. El-Laham, P. M. Djurić, and F. Hlawatsch. *Fusion of probability density functions*. Proceedings of the IEEE, 2022.
- V. Bordignon, V. Matta and A. H. Sayed. *Partial information sharing over social learning networks.* IEEE Trans. Information Theory, 2023.
- V. Bordignon, S. Vlaski, V. Matta and A. H. Sayed. *Learning from heterogeneous data based on social interactions over graphs.* IEEE Trans. Information Theory, 2023.

Concluding Remarks

There are several open questions and problems:

- New update/pooling rules
- Tracing the route of information (topology inference), privacy issues
- Optimality and performance guarantees
- Experimental analysis, proposing and testing new cognition models
- And much more…

If you are interested in further details, please send me an e-mail

vmatta@unisa.it

Thank you for attending!