Multilevel Prediction of the Voter Model with the Information Bottleneck Method

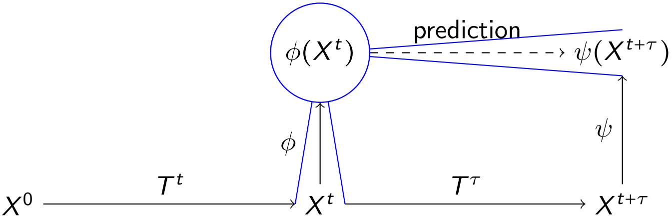
Robin Lamarche-Perrin

Max Planck Institute for Mathematics in the Sciences

General Setting: Revisiting the MatheMACS Diagram

$$\begin{tikzcd}
 & \phi(X^t) \arrow[r, dashed, "\text{prediction}"] & \psi(X^{t+\tau}) \\
X^0 \arrow[r, "T^t"] & X^t \arrow[u, "\phi"] \arrow[r, "T^\tau"] & X^{t+\tau} \arrow[u, "\psi"]
\end{tikzcd}$$

- System State: $X \in \Sigma$
- Current Time: $t \in \mathbb{N}$
- Delay: $\tau \in \mathbb{N}$
- Transitions: $T(X^{t+1}|X^t)$ (Markovian and time-homogeneous)
- Pre-measurement: stochastic map ϕ defined by $\Pr(\phi(X)|X)$
- Post-measurement: stochastic map ψ defined by $\Pr(\psi(X)|X)$



- Minimize Measurement Complexity: $min_\phi I(X^t; \phi(X^t))$
- Maximize Anticipatory Capacity: $max_\phi I(\phi(X^t); \psi(X^{t+\tau}))$

Problem

Given a time t, a delay τ, a post-measurement ψ, and a trade-off parameter $\beta \in \mathbb{R}^+$, **find** a pre-measurement ϕ with minimal complexity and maximal anticipatory capacity:

$$\min_\phi \; I(X^t; \phi(X^t)) \; - \; \beta \; I(\phi(X^t); \psi(X^{t+\tau}))$$

Constraining the Set of Feasible Measurements

$$
\begin{tikzcd}
\{\phi_1(X^t), & \phi_2(X^t), & \dots, & \phi_k(X^t)\} \arrow[r, dashed, "\text{prediction}"] & \psi(X^{t+\tau}) \\
X^0 \arrow[rrr, "T^t"] & & & X^t \arrow[lll, "\phi_1"'] \arrow[ull, "\phi_2"'] \arrow[u, "\phi_k"'] \arrow[r, "T^\tau"] & X^{t+\tau} \arrow[u, "\psi"']
\end{tikzcd}
$$

- Set of Ground Measurements: $\Phi_g = \{\phi_1, \dots, \phi_k\}$
- Feasible Measurement: $\Phi \subset \Phi_g$ with $\Phi(X) = (\phi(X))_{\phi \in \Phi}$
- and the IB problem consequently becomes the following:

$$\min_{\Phi \subset \Phi_g} \; I(X^t; \Phi(X^t)) \; - \; \beta \; I(\Phi(X^t); \psi(X^{t+\tau}))$$

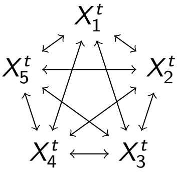
Application to the Voter Model
(Banisch & Lima, 2012)

$$
\begin{tikzcd}
& X_1^t = 1 \arrow[dl] \arrow[dr] \arrow[rrr, dashed, "(\omega_1, \omega_2)"] & & & X_1^{t+1} = 1 \arrow[dl] \arrow[dr] \arrow[rrr, dashed, "(\omega_2, \omega_3)"] & & & X_1^{t+2} = 1 \arrow[dl] \arrow[dr] & \\
X_2^t = 0 \arrow[rr, leftrightarrow] & & X_3^t = 0 & X_2^{t+1} = 1 \arrow[rr, leftrightarrow] & & X_3^{t+1} = 0 & X_2^{t+2} = 1 \arrow[rr, leftrightarrow] & & X_3^{t+2} = 1
\end{tikzcd}
$$

- Set of Agents: $\Omega = \{\omega_1, \dots, \omega_N\}$
- State of Agent ω_i: $X_i \in \{0, 1\}$
- System State: $X = (X_1, \dots, X_N) \in \Sigma = \{0, 1\}^N$
- Transitions $T(X_i^{t+1} | X^t)$ according to a weighted directed graph:

if arc (ω_i, ω_j) is selected, then $\quad X_j^{t+1} = X_i^t \quad$ and $\quad \forall k \neq j, \quad X_k^{t+1} = X_k^t$

The Case of the Complete Graph



- All arcs are equally likely: $\forall t \in \mathbb{N},\ \forall \sigma_0 \in \{0,1\}^N,\ \forall \sigma_1 \in \{0,1\}^N,$

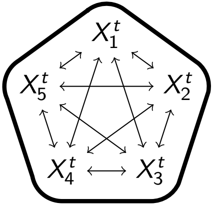
$$T\left(X^{t+1} = \sigma_1 | X^t = \sigma_0\right)\right) = \begin{cases} \frac{\overline{\sigma_0}\,(N-\overline{\sigma_0})}{N\,(N-1)} & \text{if } \overline{\sigma_1} = \overline{\sigma_0} \pm 1 \\ \frac{(N-\overline{\sigma_0})^2 + \overline{\sigma_0}^2 - N}{N\,(N-1)} & \text{if } \overline{\sigma_1} = \overline{\sigma_0} \\ 0 & \text{else} \end{cases}$$

  where $\overline{\sigma} = \sum_{s \in \sigma} s$.
- Uniformly-distributed Initial State: $\forall \sigma \in \{0,1\}^N,\ p(X^0 = \sigma) = 2^{-N}$
- Two Equiprobable Attractors: $X^\infty \in \{(0,\ldots,0),(1,\ldots,1)\}$

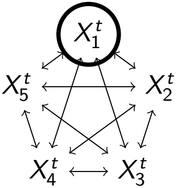
Measurement of the Voter Model

- Ground Measurements: $\forall A \subset \Omega,\ \phi_A(X) = \sum_{\omega_i \in A} X_i$

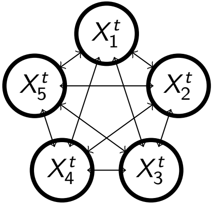
- Macro: $\phi_\top(X) = X_1 + \ldots + X_N$



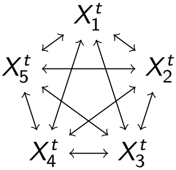
- One-agent: $\phi_{\{\omega_i\}}(X) = X_i$



- Micro: $\phi_\bot(X) = (X_1, \ldots, X_N)$



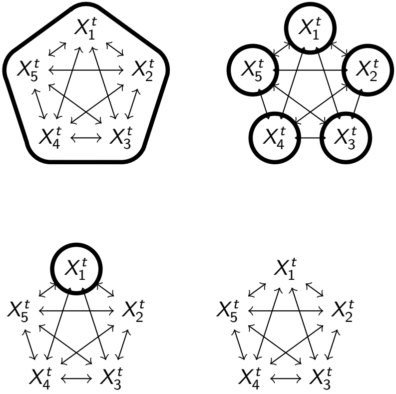
- Empty: $\phi_\varnothing(X) = 0$

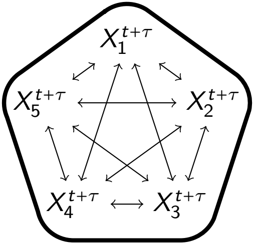


Predicting the Macroscopic Measurement

Which one of these four pre-measurements...

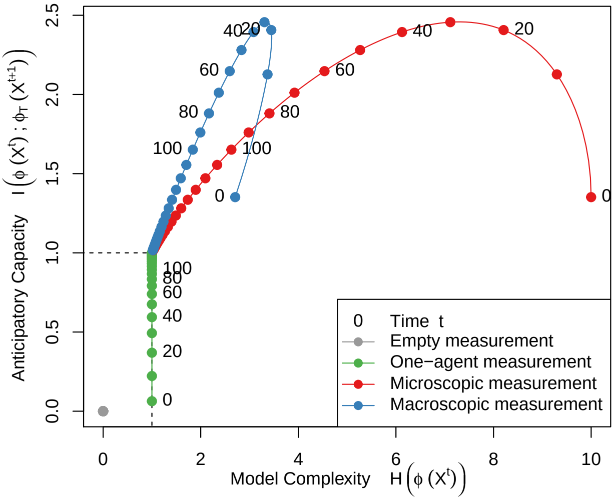
...optimally predicts the macroscopic measurement?





Predicting the Macroscopic Measurement (variable time)

- Size $N = 10$, delay $\tau = 1$, and variable time $t \in \mathbb{N}$



Predicting the Macroscopic Measurement (variable time)

1. At any time, the macroscopic measurement has a lower complexity than the microscopic measurement:

$$\forall t \in \mathbb{N}, \quad H(\phi_\top(X^t)) < H(\phi_\perp(X^t)).$$

2. At any time and for any delay, the microscopic measurement and the macroscopic measurements have the same anticipatory capacity:

$$\forall t \in \mathbb{N},\ \forall \tau \in \mathbb{N}, \quad I(\phi_\perp(X^t); \phi_\top(X^{t+\tau})) = I(\phi_\top(X^t); \phi_\top(X^{t+\tau})).$$

3. For any delay, as time goes by, the complexity and anticipatory capacity of any non-empty measurement converge to those of a fair Bernoulli variable predicting itself:

$$\forall \phi \neq \phi_\varnothing,\ \forall \tau \in \mathbb{N}, \quad H(\phi(X^t)) \xrightarrow[t\to\infty]{} 1 \quad \text{and} \quad I(\phi(X^t); \phi_\top(X^{t+\tau})) \xrightarrow[t\to\infty]{} 1.$$

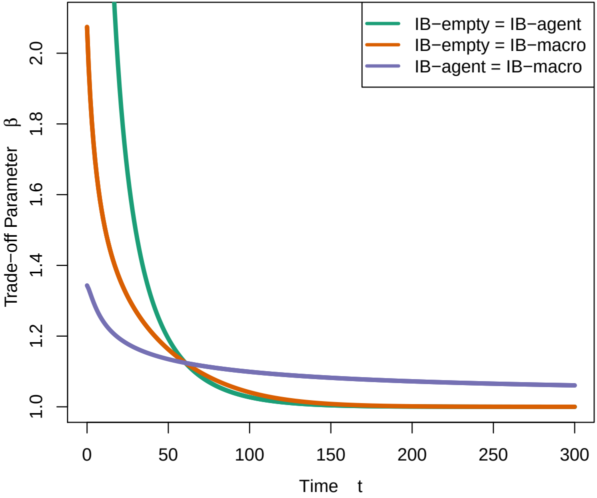
Optimal-Bottleneck Diagrams

- Parameters: $(t, \tau, \beta) \in \mathbb{N} \times \mathbb{N} \times \mathbb{R}^+$
- IB-equivalence: $B^{t,\tau}_{\phi_1,\phi_2} = \{\beta \in \mathbb{R}^+ \setminus IB^{t,\tau}_\beta(\phi_1;\psi) = IB^{t,\tau}_\beta(\phi_2;\psi)\}$

	Border	IB-diagram
$(H_1 > H_2 \wedge I_1 > I_2) \vee (H_1 < H_2 \wedge I_1 < I_2) \Rightarrow \lvert B^{t,\tau}_{\phi_1,\phi_2} \rvert = 1$	$\beta^{t,\tau}_{\phi_1,\phi_2} = \frac{H_2 - H_1}{I_2 - I_1} > 0$	Two regions
$(H_1 \neq H_2 \wedge I_1 = I_2) \Rightarrow B^{t,\tau}_{\phi_1,\phi_2} = \varnothing$	$\beta^{t,\tau}_{\phi_1,\phi_2} = +\infty$	Sup. border
$(H_1 = H_2 \wedge I_1 \neq I_2) \Rightarrow \lvert B^{t,\tau}_{\phi_1,\phi_2} \rvert = 1$	$\beta^{t,\tau}_{\phi_1,\phi_2} = 0$	Inf. border
$(H_1 = H_2 \wedge I_1 = I_2) \Rightarrow B^{t,\tau}_{\phi_1,\phi_2} = \mathbb{R}^+$	Defined everywhere	No region
$(H_1 > H_2 \wedge I_1 < I_2) \vee (H_1 < H_2 \wedge I_1 > I_2) \Rightarrow B^{t,\tau}_{\phi_1,\phi_2} = \varnothing$	Defined nowhere	One region

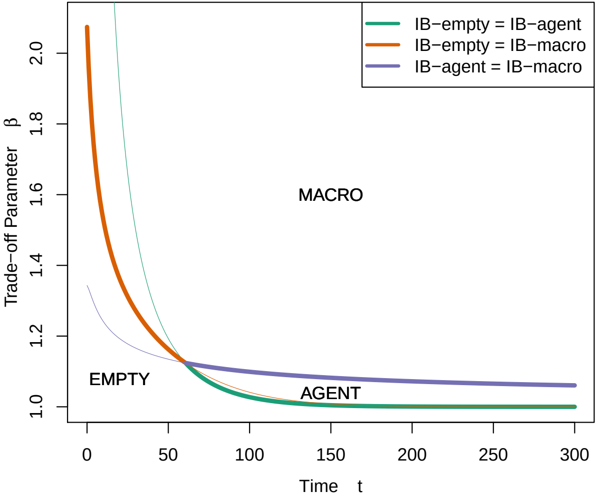
Predicting the Macroscopic Measurement (variable time)

- Size $N = 10$, delay $\tau = 1$, and variable time $t \in \mathbb{N}$



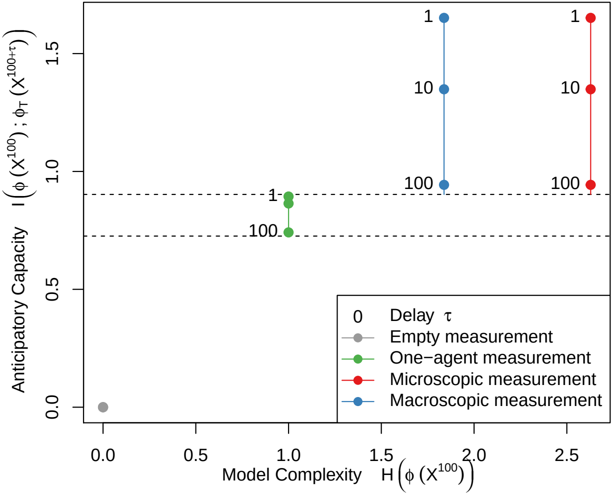
Predicting the Macroscopic Measurement (variable time)

- Size $N = 10$, delay $\tau = 1$, and variable time $t \in \mathbb{N}$



Predicting the Macroscopic Measurement (variable delay)

- Size $N = 10$, time $t = 100$, and variable delay $\tau \in \mathbb{N}$



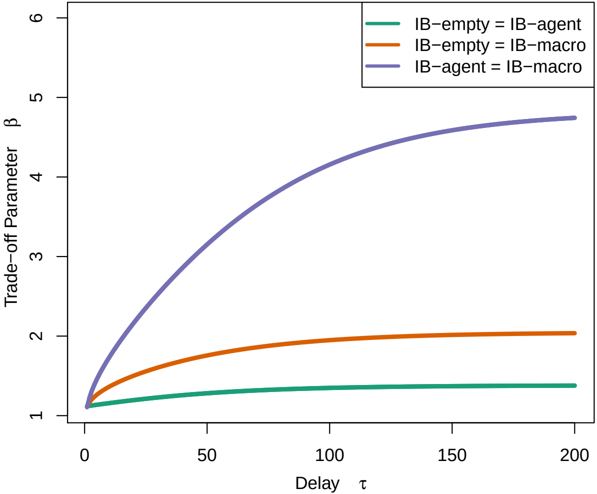
Predicting the Macroscopic Measurement (variable delay)

4. At any time, as the delay increases, the anticipatory capacity of any non-empty measurement converges to a non-null value:

$$\forall \phi \neq \phi_{\varnothing},\ \forall t \in \mathbb{N}, \quad \lim_{\tau \to \infty} I(\phi(X^t); \phi_\top(X^{t+\tau})) > 0.$$

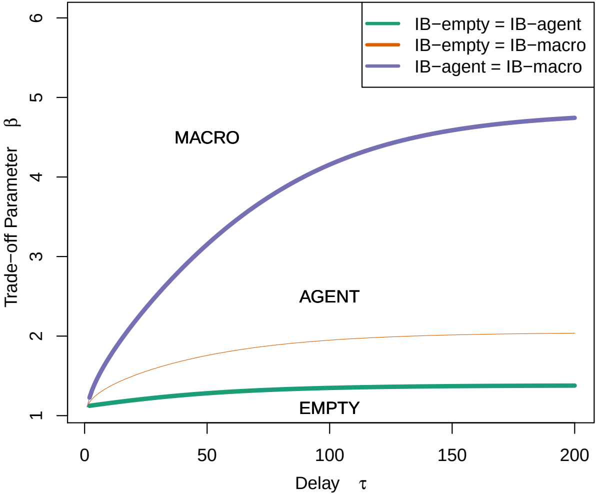
Predicting the Macroscopic Measurement (variable delay)

- Size $N = 10$, time $t = 100$, and variable delay $\tau \in \mathbb{N}$



Predicting the Macroscopic Measurement (variable delay)

- Size $N = 10$, time $t = 100$, and variable delay $\tau \in \mathbb{N}$



Predicting the Macroscopic Measurement

5. At any time and for any delay, the border between the microscopic measurement and the macroscopic measurement is infinite:

$$\forall t \in \mathbb{N},\ \forall \tau \in \mathbb{N}, \quad \beta_{\top,\perp}^{t,\tau} = +\infty.$$

Hence, for any finite value of the trade-off parameter, the macroscopic measurement is better than the microscopic measurement.

6. At any time and for any delay, the border between the empty measurement and any other measurement is finite, it converges to 1 from above as time goes by, and to a finite value from below as delay increases:

$$\forall \phi \neq \phi_\varnothing, \quad \begin{aligned} &\forall t \in \mathbb{N},\ \forall \tau \in \mathbb{N},\ \beta_{\varnothing,\phi}^{t,\tau} \in \,]1, +\infty[, \\ &\forall \tau \in \mathbb{N},\ \beta_{\varnothing,\phi}^{t,\tau} \xrightarrow[t \to \infty]{} 1, \\ &\forall t \in \mathbb{N},\ \beta_{\varnothing,\phi}^{t,\tau} \xrightarrow[\tau \to \infty]{} \beta_{\varnothing,\phi}^{t,\infty} \in \,]1, +\infty[. \end{aligned}$$

Hence, for $\beta \in [0, 1]$, the empty measurement is always better than any other measurement and, for $\beta \in \,]1, +\infty[$, there is a time after which any measurement becomes better than the empty measurement. Moreover, for $\beta \in [0, \beta_{\varnothing,\phi}^{t,\infty}]$, the empty measurement is better than ϕ to predict the final state and, for $\beta \in]\beta_{\varnothing,\phi}^{t,\infty}, +\infty[$, ϕ is better than the empty measurement to predict the final state.

Predicting the Macroscopic Measurement

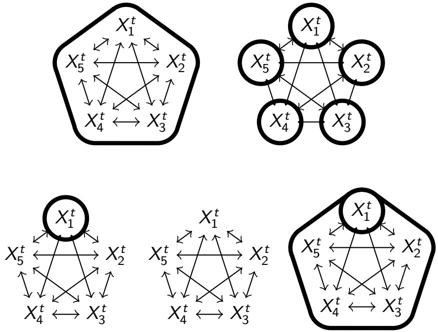
7. At any time and for any delay, the border between the macroscopic measurement and the one-agent measurement is strictly positive:

$$\forall \omega \in \Omega,\ \forall t \in \mathbb{N},\ \forall \tau \in \mathbb{N}, \quad \beta^{t,\tau}_{\{\omega\},\top} \in \left]0, +\infty\right[.$$

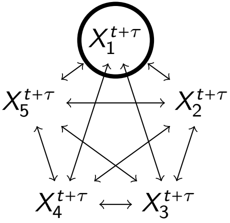
8. There is a time $t' \in \mathbb{N}$ such that $\forall t < t'$, $\beta^{t,\tau}_{\{\omega\},\top} < \beta^{t,\tau}_{\varnothing,\top} < \beta^{t,\tau}_{\varnothing,\{\omega\}}$, meaning that the one-agent measurement is worst than both the macroscopic and the empty measurements and, $\forall t > t_1$, $\beta^{t,\tau}_{\varnothing,\{\omega\}} < \beta^{t,\tau}_{\varnothing,\top} < \beta^{t,\tau}_{\{\omega\},\top}$, meaning that there is a range $\left]\beta^{t}_{\varnothing,\{\omega\}}, \beta^{t}_{\{\omega\},\top}\right[$ of values of the trade-off parameter for which the one-agent measurement is better than both the macroscopic and the empty measurements.
9. Since $\beta^{t,\tau}_{\{\omega\},\top} \xrightarrow[t\to\infty]{} 1$, this range converges to the empty interval as time goes by, meaning that, for $\beta \in \left]1, +\infty\right[$, the macroscopic measurement becomes better than the empty measurement after a certain time.
10. Since $\beta^{t,\tau}_{\{\omega\},\top}$ converges to a finite value when $\tau \to \infty$, there is a value of the trade-of parameter above which the macroscopic measurement is always better than the one-agent measurement.

Predicting the One-agent Measurement

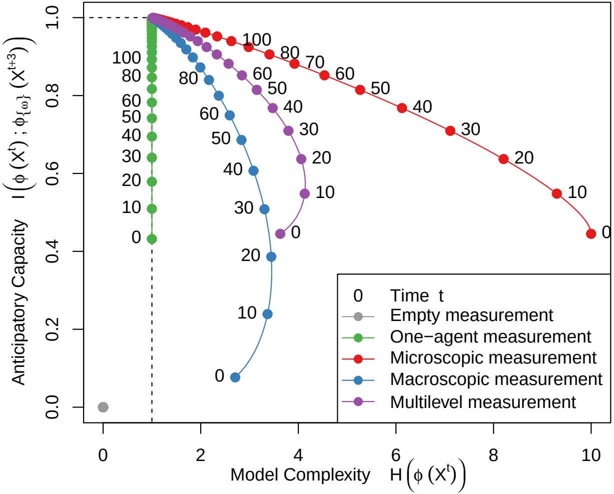
Which one of these five pre-measurements...



...optimally predicts the one-agent measurement?

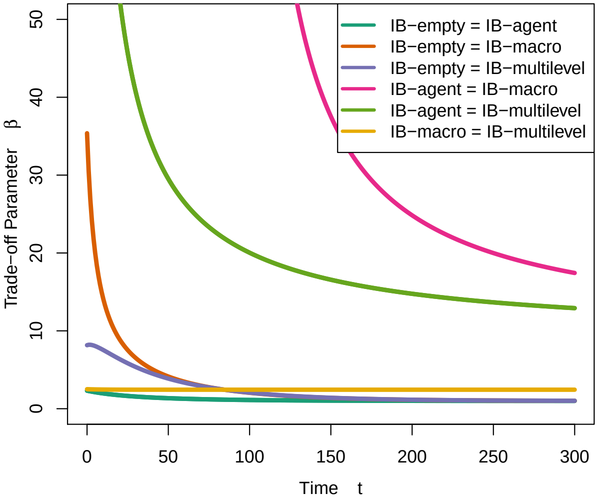


- Size $N = 10$, delay $\tau = 3$, and variable time $t \in \mathbb{N}$

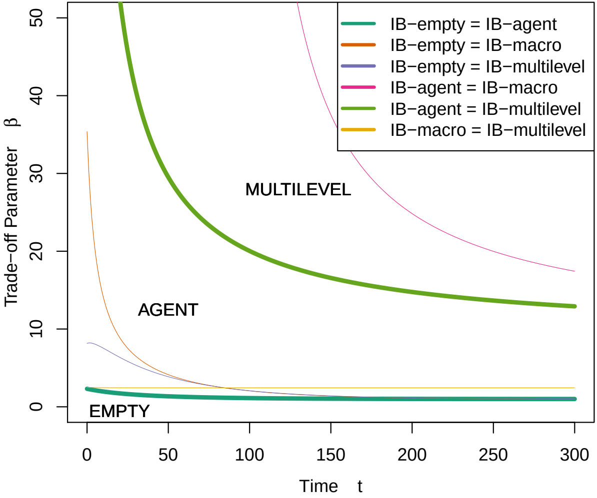


Predicting the One-agent Measurement (variable time)

- Size $N = 10$, delay $\tau = 3$, and variable time $t \in \mathbb{N}$

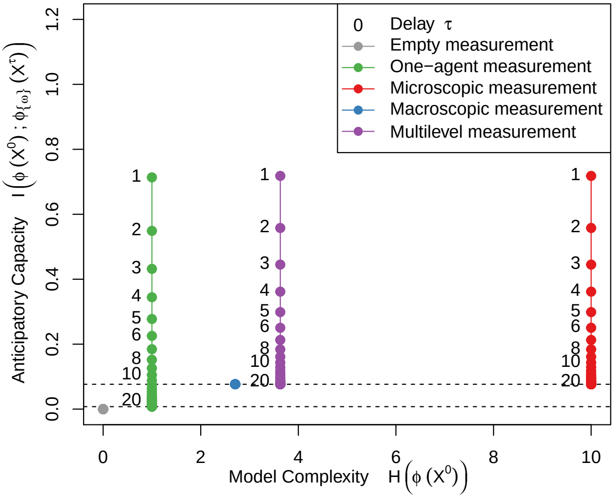


- Size $N = 10$, delay $\tau = 3$, and variable time $t \in \mathbb{N}$



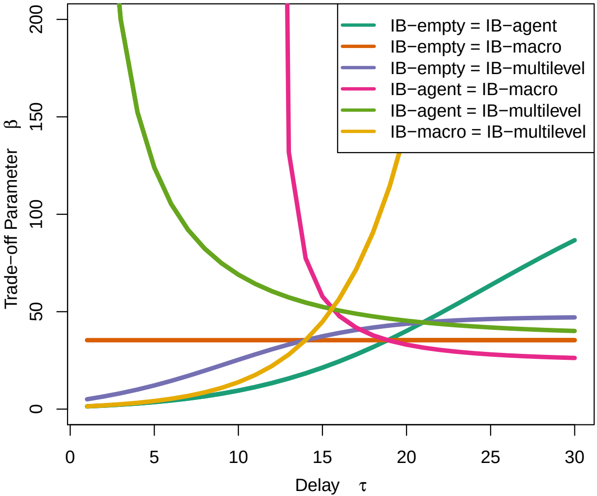
Predicting the One-agent Measurement (variable delay)

- Size $N = 10$, time $t = 0$, and variable delay $\tau \in \mathbb{N}$



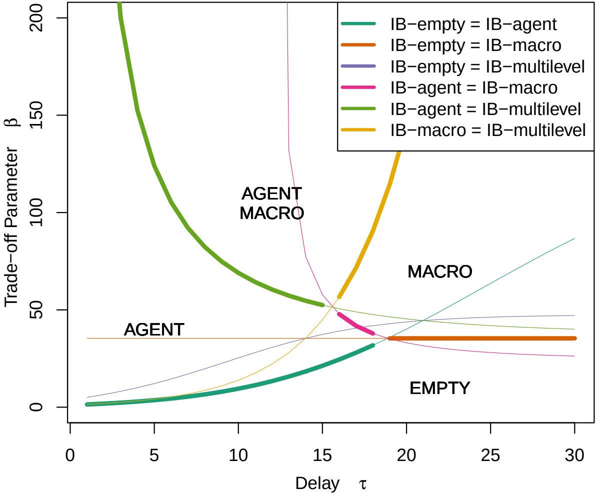
Predicting the One-agent Measurement (variable delay)

- Size $N = 10$, time $t = 0$, and variable delay $\tau \in \mathbb{N}$



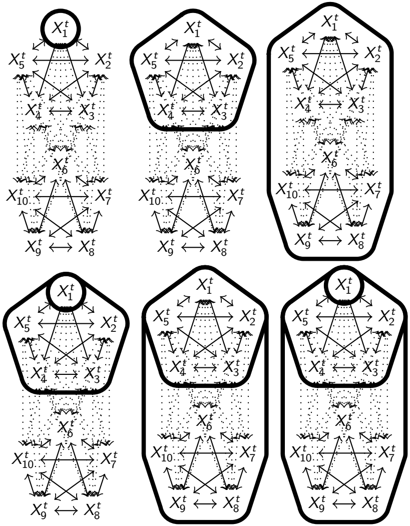
Predicting the One-agent Measurement (variable delay)

- Size $N = 10$, time $t = 0$, and variable delay $\tau \in \mathbb{N}$

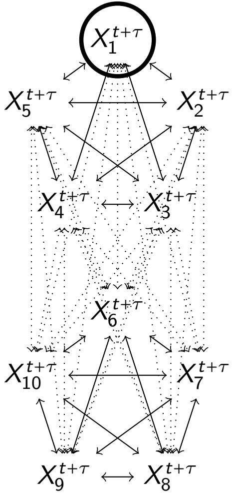


The Two-communities Case

Which one of these six pre-measurements...

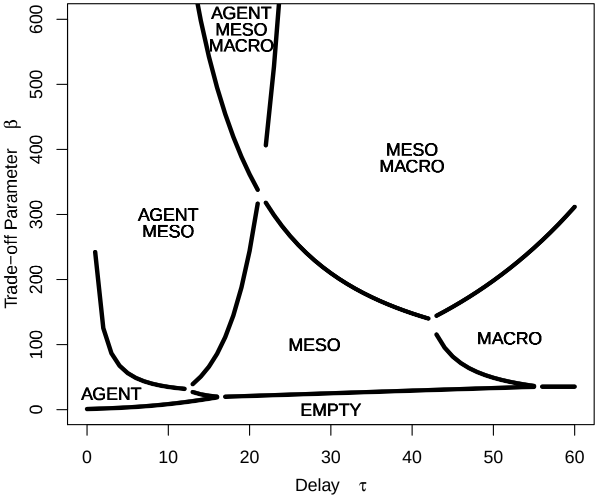


...optimally predicts the one-agent measurement?



Predicting the One-agent Measurement (variable delay)

- Size $(N_1, N_2) = (5, 5)$, inter-rate $\rho = 0.1$, time $t = 0$, and variable delay $\tau \in \mathbb{N}$



Tractability of the IB Optimization Problem

- NP-complete in the general case.
- Which additional assumptions regarding the set of ground measurements $\Phi_g = \{\phi_1, \dots, \phi_k\}$ are sufficient for tractability?
- Example:

Definition

Given a post-measurement ψ, the IB trade-off is additively decomposable *if, for all feasible measurement $\Phi \subset \Phi_g$ and for all parameters $(t, \tau, \beta) \in \mathbb{N} \times \mathbb{N} \times \mathbb{R}^+$, we have:*

$$IB_\beta^{t,\tau}(\Phi; \psi) = \sum_{\phi \in \Phi} IB_\beta^{t,\tau}(\phi, \psi)$$

Theorem 1

If the IB trade-off is additively decomposable, then, for all $\beta \in \mathbb{R}^+$, $\Phi_\beta = \{\phi \in \Phi_g \setminus IB_\beta^{t,\tau}(\phi; \psi) < 0\}$ is an optimal measurement for the trade-off parameter β.

Conditions for Additive Decomposability

Theorem 2

The IB trade-off is additively decomposable if and only if the ground measurements are mutually independent per se, given the current state, and given the predicted state.
Formally, that is when, for all $(\phi_1, \phi_2) \in \Phi_g^2$, we have:

$$
\begin{array}{lllll}
\textit{(C1)} & I(\phi_1(X^t); \phi_2(X^t)) & = & 0 \\
\textit{(C2)} & I(\phi_1(X^t); \phi_2(X^t) | X^t) & = & 0 \\
\textit{(C3)} & I(\phi_1(X^t); \phi_2(X^t) | \psi(X^{t+\tau})) & = & 0
\end{array}
$$

Problem

Assuming that $\exists \epsilon \in \mathbb{R}^+$ *such that*

$$
\begin{array}{lll}
(C1b) & I(\phi_1(X^t); \phi_2(X^t)) & < \epsilon \\
(C2b) & I(\phi_1(X^t); \phi_2(X^t) | X^t) & < \epsilon \\
(C3b) & I(\phi_1(X^t); \phi_2(X^t) | \psi(X^{t+\tau})) & < \epsilon
\end{array}
$$

is there $\eta \in \mathbb{R}^+$ *such that*

$$\forall \Phi \subset \Phi_g, \quad IB_\beta^{t,\tau}(\Phi_\beta; \psi) \leq IB_\beta^{t,\tau}(\Phi; \psi) + \eta$$

for any instance of the problem. In other terms, can we guarantee provable optimality bounds when tractability conditions are slightly relaxed? If so, how can one derives such bounds from ϵ *and* β*?*

Application to Multivariate Systems

$\forall A \subset \Omega$, we also mark $A^t = \pi_A(X^t)$.

Theorem 3

For any two subsets $A_1 \subset \Omega$ and $A_2 \subset \Omega$, we have $I(A_1^t; A_2^t | X^t) = 0$ (C2).

Theorem 4

For any two subsets $A_1 \subset \Omega$ and $A_2 \subset \Omega$, we have $I(A_1^t; A_2^t) = 0$ (C1) if and only if $H(A_1^t \cap A_2^t) = 0$ and $I(A_i^t \backslash A_j^t; A_j^t \backslash A_i^t) = 0$. This is the case in particular when X^t is uniformly distributed over S^N and when $A_1 \cap A_2 = \varnothing$.

Theorem 5

For any two subsets $A_1 \subset \Omega$ and $A_2 \subset \Omega$ and for any other two subsets $B_1 \subset \Omega$ and $B_2 \subset \Omega$, we have $I(A_1^t; A_2^t | B_1^{t+\tau}, B_2^{t+\tau}) = 0$ if and only if

- $I(A_1^t; A_2^t) = 0$,
- $I(B_1^{t+\tau}; B_2^{t+\tau}) = 0$,
- $I(A_1^t; B_2^{t+\tau}) = 0$ *and* $I(A_2^t; B_1^{t+\tau}) = 0$ *(or the converse).*

Thank you for your attention