Inverse Problems for Philosophers

Bridging the gap between agent-based models and behavioral data

Lucas Gautheron[1,2] (Ph.D. candidate)

[1]Interdisciplinary Center for Science and Technology Studies, Wuppertal, Germany
[2]Département d'Études Cognitives, École Normale Supérieure, Paris, France

University of Bochum, January 2025

Summary

1. Inverse problems for philosophers and agent-based modelers

2. A case-study of conventions: the metric signature in particle physics
   - How do physicists choose which convention to use in their own papers?
   - How do scientists resolve conflicting preferences in collaborations?
   - How do physicists' preferences get formed?

Why should philosophers care about data?

Reasons are:

- **Intellectual**: assess whether models capture what is actually going on in situations of interest / use data as a source of inspiration

Why should philosophers care about data?

Reasons are:

- **Intellectual**: assess whether models capture what is actually going on in situations of interest / use data as a source of inspiration
- **Methodological**: Non-empirical validation is fallible. Example: “robustness” (insensitivity to model assumptions/parameters) $\Rightarrow$ what if the outcome really is contingent on certain circumstances (the values of underlying parameters, the topology of some relevant network, etc.)

Why should philosophers care about data?

Reasons are:

- **Intellectual**: assess whether models capture what is actually going on in situations of interest / use data as a source of inspiration
- **Methodological**: Non-empirical validation is fallible. Example: "robustness" (insensitivity to model assumptions/parameters) $\Rightarrow$ what if the outcome really is contingent on certain circumstances (the values of underlying parameters, the topology of some relevant network, etc.)
- **Practical**: normative insights from models without connection to data may not be translatable into interventions/policies (abstract parameters in a computational model do not immediately connect to actionable parameters!)

Summary

1. Inverse problems for philosophers and agent-based modelers

2. A case-study of conventions: the metric signature in particle physics
   - How do physicists choose which convention to use in their own papers?
   - How do scientists resolve conflicting preferences in collaborations?
   - How do physicists' preferences get formed?

What are inverse problems

- Inverse problems seek to **infer the invisible causes underlying a set of observations**.

What are inverse problems

- Inverse problems seek to **infer the invisible causes underlying a set of observations**.
- In the context of Agent-Based Modeling:

Rules governing agents' behavior $\xrightarrow{\text{“Forward problem”}}$ Outcome of agents' behavior

What are inverse problems

- Inverse problems seek to **infer the invisible causes underlying a set of observations**.
- In the context of Agent-Based Modeling:

$$\text{Rules governing agents' behavior} \; \underset{\text{``Inverse problem''}}{\overset{\text{``Forward problem''}}{\rightleftarrows}} \; \text{Outcome of agents' behavior}$$

What are inverse problems

- Inverse problems seek to **infer the invisible causes underlying a set of observations**.
- In the context of Agent-Based Modeling:

Rules governing agents' behavior $\xrightarrow{\text{“Forward problem”}}$ Outcome of agents' behavior

Rules governing agents' behavior $\xleftarrow[\text{“Inverse problem”}]{}$ Outcome of agents' behavior

- Inverse problems are **hard**:

What are inverse problems

- Inverse problems seek to **infer the invisible causes underlying a set of observations**.
- In the context of Agent-Based Modeling:

$$\text{Rules governing agents' behavior} \quad \underset{\text{"Inverse problem"}}{\overset{\text{"Forward problem"}}{\rightleftarrows}} \quad \text{Outcome of agents' behavior}$$

- Inverse problems are **hard**:
  1. **Identifiability problems** (underdetermination): many causes could have produced a given outcome

What are inverse problems

- Inverse problems seek to **infer the invisible causes underlying a set of observations**.
- In the context of Agent-Based Modeling:

Rules governing agents' behavior	"Forward problem" $\longrightarrow$ $\longleftarrow$ "Inverse problem"	Outcome of agents' behavior

- Inverse problems are **hard**:
  1. **Identifiability problems** (underdetermination): many causes could have produced a given outcome
  2. **Misspecification problems**: inverse problems may produce misleading results when modeling assumptions are "too wrong".

What are inverse problems

- Inverse problems seek to **infer the invisible causes underlying a set of observations**.
- In the context of Agent-Based Modeling:

Rules governing agents' behavior	"Forward problem" → ← "Inverse problem"	Outcome of agents' behavior

- Inverse problems are **hard**:
  1. **Identifiability problems** (underdetermination): many causes could have produced a given outcome
  2. **Misspecification problems**: inverse problems may produce misleading results when modeling assumptions are "too wrong".
  3. **Computational problems**: solving inverse problems often involves intractable computations and requires approximation schemes.

Bayesian inference for inverse problems

- Both forward models and inverse problems have a stochastic/probabilistic component (random initialization, partially random decisions, uncertainty quantification...)
- We appeal to **probabilities** and **Bayesian inference**.

$$\begin{array}{ccc} \text{Rules governing} & \text{“Forward problem”} & \text{Outcome of} \\ \text{agents' behavior} & \text{Estimate } \boldsymbol{P(O|M)} & \text{agents' behavior} \\ \textbf{Model } \boldsymbol{M} & \xrightarrow{\hspace{6em}} & \textbf{Outcome } \boldsymbol{O} \\ & \xleftarrow{\hspace{6em}} & \\ & \text{“Inverse problem”} & \\ & \text{Estimate } \boldsymbol{P(M|O)} & \end{array}$$

$$P(M|O) = \frac{P(O|M)\,\overbrace{P(M)}^{\text{Prior}}}{P(O)} \tag{1}$$

Model comparison and parameter estimation

$$P(M|O) = \frac{P(O|M)\overbrace{P(M)}^{\text{Prior}}}{P(O)} \tag{2}$$

Model comparison and parameter estimation

$$P(\theta|O) = \frac{P(O|\theta)\overbrace{P(\theta)}^{\text{Prior}}}{P(O)} \tag{2}$$

Summary

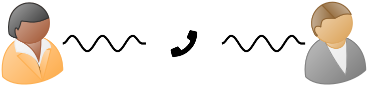
1 Inverse problems for philosophers and agent-based modelers

2 A case-study of conventions: the metric signature in particle physics

- How do physicists choose which convention to use in their own papers?
- How do scientists resolve conflicting preferences in collaborations?
- How do physicists' preferences get formed?

Conventions

- **Coordination problems** arise when individuals would benefit from acting in a mutually compatible way, but it is somehow non-trivial to do so (Lewis, 1969).



Conventions

- **Coordination problems** arise when individuals would benefit from acting in a mutually compatible way, but it is somehow non-trivial to do so (Lewis, 1969).

Conventions

- **Coordination problems** arise when individuals would benefit from acting in a mutually compatible way, but it is somehow non-trivial to do so (Lewis, 1969).

	Bob calls back	**Bob awaits**
Alice calls back	0,0	1,1
Alice awaits	1,1	0,0

Conventions

- **Coordination problems** arise when individuals would benefit from acting in a mutually compatible way, but it is somehow non-trivial to do so (Lewis, 1969).

	Bob calls back	**Bob awaits**
Alice calls back	0,0	1,1
Alice awaits	1,1	0,0

- **"Conventions"** are cultural tools for solving coordination problems by providing individuals with expectations about how others will behave. These expectations suggest particular courses of action.

Conventions

- **Coordination problems** arise when individuals would benefit from acting in a mutually compatible way, but it is somehow non-trivial to do so (Lewis, 1969).

	Bob calls back	**Bob awaits**
Alice calls back	0,0	1,1
Alice awaits	1,1	0,0

- **"Conventions"** are cultural tools for solving coordination problems by providing individuals with expectations about how others will behave. These expectations suggest particular courses of action.
  - Example: left-hand or right-hand traffic.

- **Coordination problems** arise when individuals would benefit from acting in a mutually compatible way, but it is somehow non-trivial to do so (Lewis, 1969).

	Bob calls back	**Bob awaits**
Alice calls back	0,0	1,1
Alice awaits	1,1	0,0

- **“Conventions”** are cultural tools for solving coordination problems by providing individuals with expectations about how others will behave. These expectations suggest particular courses of action.
  - Example: left-hand or right-hand traffic.
  - Language! “The syllable ‘big’ could have meant ‘small’ for all we care, and the red light could have meant ‘go”’ (Quine, foreword to Lewis 1969)

Conventions in the literature

- Generally based on Lewis: conventions as solutions to coordination games (Lewis, 1969)

Conventions in the literature

- Generally based on Lewis: conventions as solutions to coordination games (Lewis, 1969)
    1. Can conventions emerge spontaneously from dyadic interactions alone? (Centola and Baronchelli, 2015; Hawkins, Goodman, and Goldstone, 2019)

Conventions in the literature

- Generally based on Lewis: conventions as solutions to coordination games (Lewis, 1969)
  1. Can conventions emerge spontaneously from dyadic interactions alone? (Centola and Baronchelli, 2015; Hawkins, Goodman, and Goldstone, 2019)
  2. How does the topology of social networks influence the propagation of conventions via dyadic interactions? (Pujol et al., 2005; Delgado, 2002)

Conventions in the literature

- Generally based on Lewis: conventions as solutions to coordination games (Lewis, 1969)
  1. Can conventions emerge spontaneously from dyadic interactions alone? (Centola and Baronchelli, 2015; Hawkins, Goodman, and Goldstone, 2019)
  2. How does the topology of social networks influence the propagation of conventions via dyadic interactions? (Pujol et al., 2005; Delgado, 2002)
  3. How to measure the degree of conventionality of a convention? (O'Connor, 2020)

Conventions in the literature

- Generally based on Lewis: conventions as solutions to coordination games (Lewis, 1969)
  1. Can conventions emerge spontaneously from dyadic interactions alone? (Centola and Baronchelli, 2015; Hawkins, Goodman, and Goldstone, 2019)
  2. How does the topology of social networks influence the propagation of conventions via dyadic interactions? (Pujol et al., 2005; Delgado, 2002)
  3. How to measure the degree of conventionality of a convention? (O'Connor, 2020)

  Most often: idealized formal models or controlled experiments. Few studies in naturalistic settings!

A case-study from high-energy physics

- Relativistic theories: unified description of spacetime.

A case-study from high-energy physics

- Relativistic theories: unified description of spacetime.
- The metric tensor ($g_{\mu\nu}$) captures the metric properties of spacetime; e.g. the pseudo- distance between events (t_1, x_1, y_1, z_1) and (t_2, x_2, y_2, z_2). **Two possible descriptions** (metric signatures):

A case-study from high-energy physics

- Relativistic theories: unified description of spacetime.
- The metric tensor ($g_{\mu\nu}$) captures the metric properties of spacetime; e.g. the pseudo- distance between events (t_1, x_1, y_1, z_1) and (t_2, x_2, y_2, z_2). **Two possible descriptions** (metric signatures):

$$\begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \text{ or } \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{pmatrix}?$$

A case-study from high-energy physics

- Relativistic theories: unified description of spacetime.
- The metric tensor ($g_{\mu\nu}$) captures the metric properties of spacetime; e.g. the pseudo- distance between events (t_1, x_1, y_1, z_1) and (t_2, x_2, y_2, z_2). **Two possible descriptions** (metric signatures):

$$\begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \text{ or } \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{pmatrix} ?$$

$$\text{“mostly minus” (-1) or “mostly plus” } (+1) \tag{3}$$

A case-study from high-energy physics

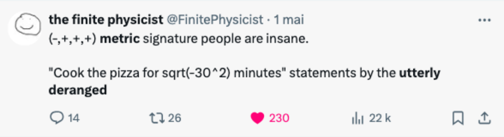
- Relativistic theories: unified description of spacetime.
- The metric tensor ($g_{\mu\nu}$) captures the metric properties of spacetime; e.g. the pseudo- distance between events (t_1, x_1, y_1, z_1) and (t_2, x_2, y_2, z_2). **Two possible descriptions** (metric signatures):

$$\begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \text{ or } \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{pmatrix}?$$

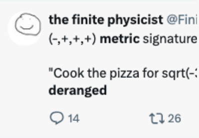
$$\text{“mostly minus” (-1) or “mostly plus” } (+1) \tag{3}$$

- Both choices are legitimate, as long as one remains consistent.

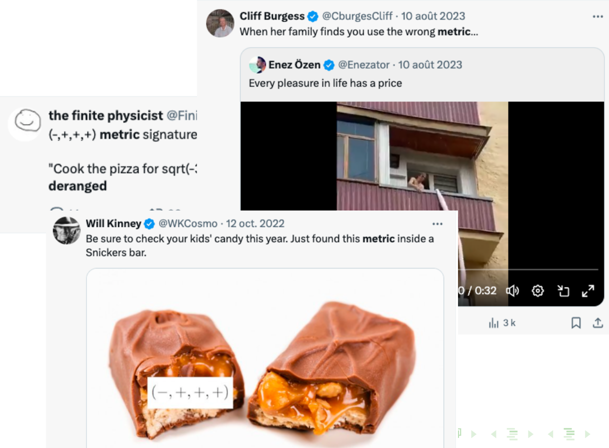
A heated debate



A heated debate

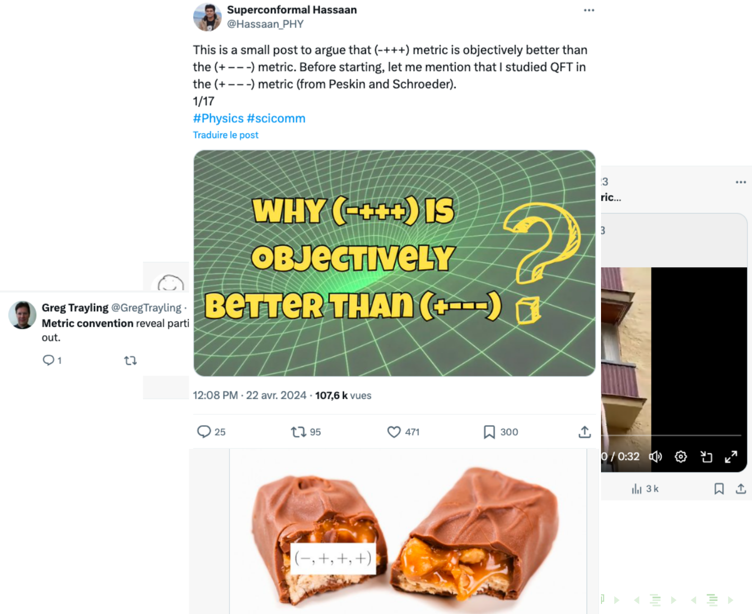


A heated debate



A heated debate

A heated debate



Inverse problems and conventions

- Let's use inverse problems to infer:

Inverse problems and conventions

- Let's use inverse problems to infer:
    1. How do scientists decide which convention to use in a paper?

Inverse problems and conventions

- Let's use inverse problems to infer:
  1. How do scientists decide which convention to use in a paper?
  2. How do they resolve conflicting preferences in collaborations?

Inverse problems and conventions

- Let's use inverse problems to infer:
  1. How do scientists decide which convention to use in a paper?
  2. How do they resolve conflicting preferences in collaborations?
  3. What factors shape scientists' preferences?

Data

- Data collected from **Inspire HEP** (authorship/citation metadata) and **arXiv** (LaTeX source)
- Categories: hep-th (high-energy physics theory), hep-ph (phenomenology), gr-qc (gravitation and cosmology), astro-ph (astrophysics)
- 22 500 papers classified according to their metric signature (mostly plus or mostly minus) using regular expressions.

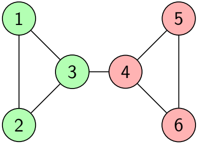
1. Inverse problems for philosophers and agent-based modelers

2. A case-study of conventions: the metric signature in particle physics
   - How do physicists choose which convention to use in their own papers?
   - How do scientists resolve conflicting preferences in collaborations?
   - How do physicists' preferences get formed?

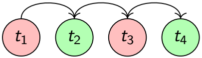
How do physicists choose which convention to use in their own papers?

Individuals' attitude towards a convention may be shaped by:

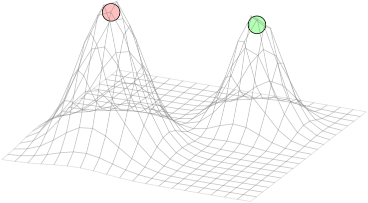
Social consistency
(coordination costs)



Sequential consistency
(switching costs)



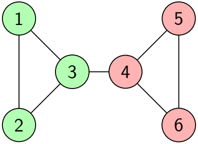
Contextual consistency
(maladaptation costs)



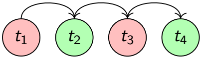
How do physicists choose which convention to use in their own papers?

Individuals' attitude towards a convention may be shaped by:

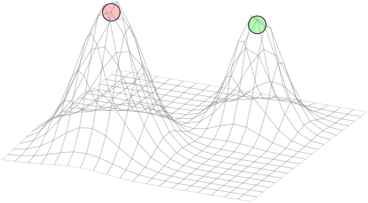
Social consistency
(coordination costs)



Sequential consistency
(switching costs)



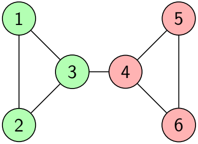
Contextual consistency
(maladaptation costs)



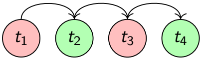
How do physicists choose which convention to use in their own papers?

Individuals' attitude towards a convention may be shaped by:

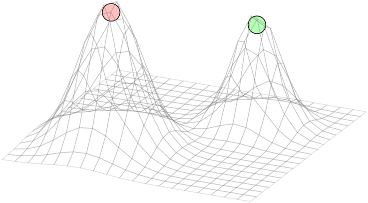
Social consistency
(coordination costs)



Sequential consistency
(switching costs)



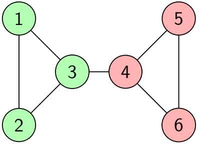
Contextual consistency
(maladaptation costs)



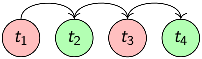
How do physicists choose which convention to use in their own papers?

Individuals' attitude towards a convention may be shaped by:

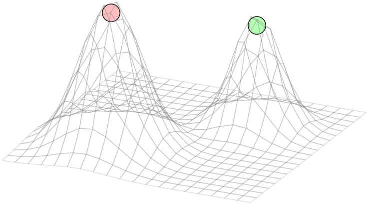
Social consistency
(coordination costs)



Sequential consistency
(switching costs)



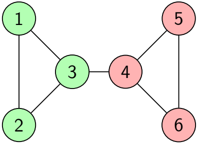
Contextual consistency
(maladaptation costs)



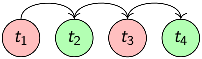
How do physicists choose which convention to use in their own papers?

Individuals' attitude towards a convention may be shaped by:

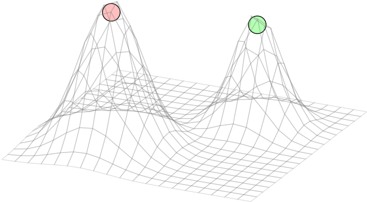
Social consistency
(coordination costs)



Sequential consistency
(switching costs)



Contextual consistency
(maladaptation costs)

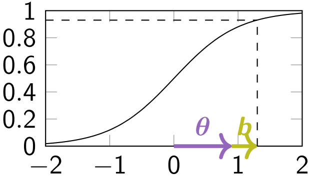


$\Rightarrow$ Are these involved in the context of the metric signature?

Sequential and contextual consistency

- publishes d in category $c_d \in \{\text{phenomenology, theory, ...}\}$. What determines which convention she uses? Our assumption:

$$P(\sigma_d = +1 | a_d = \;, c_d) = f(\underbrace{\theta(\;)}_{\text{Author's preference}} + \underbrace{b(c_d)}_{\text{Effect of research area } c_d})$$



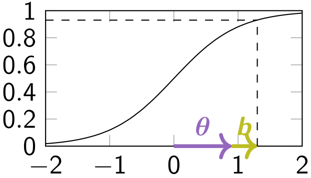
Sequential and contextual consistency

-  publishes d in category $c_d \in \{\text{phenomenology, theory, ...}\}$. What determines which convention she uses? Our assumption:

$P(\sigma_d = +1 | a_d =$  $, c_d) = f\big(\theta($  $) + b(c_d)\big)$

Author's preference

Effect of research area c_d

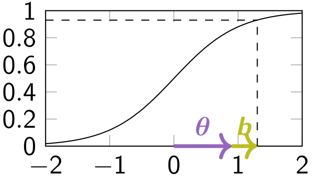


- $\theta(i) = \pm\mu$ is a latent (unobserved) parameter measuring the preference of each author i. $\theta(i) > 0$ indicates a preference for the mostly plus signature

Sequential and contextual consistency

- publishes d in category $c_d \in \{\text{phenomenology, theory, ...}\}$. What determines which convention she uses? Our assumption:

$$P(\sigma_d = +1 | a_d = \;, c_d) = f(\underbrace{\theta(\;)}_{\text{Author's preference}} + \underbrace{b(c_d)}_{\text{Effect of research area } c_d})$$



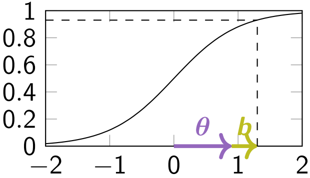
- $\theta(i) = \pm\mu$ is a latent (unobserved) parameter measuring the preference of each author i. $\theta(i) > 0$ indicates a preference for the mostly plus signature
- b_c is the unobserved bias associated with research area c

Sequential and contextual consistency

-  publishes d in category $c_d \in \{\text{phenomenology, theory, ...}\}$. What determines which convention she uses? Our assumption:

$P(\sigma_d = +1 | a_d =$  $, c_d) = f(\theta($  $) + b(c_d))$

Author's preference — Effect of research area c_d

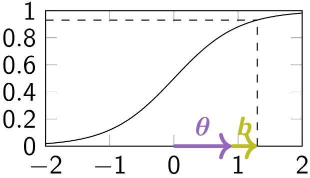


- $\theta(i) = \pm\mu$ is a latent (unobserved) parameter measuring the preference of each author i. $\theta(i) > 0$ indicates a preference for the mostly plus signature
- b_c is the unobserved bias associated with research area c
- If $|\theta| \gg |b|$, individual preferences dominate the need to adapt to a given research area

Sequential and contextual consistency

- publishes d in category $c_d \in \{\text{phenomenology, theory, ...}\}$. What determines which convention she uses? Our assumption:

$$P(\sigma_d = +1 | a_d = \quad, c_d) = f(\underbrace{\theta(\quad)}_{\text{Author's preference}} + \underbrace{b(c_d)}_{\text{Effect of research area } c_d})$$



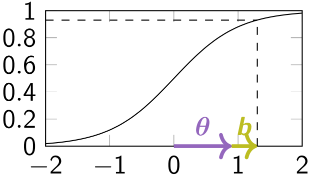
- $\theta(i) = \pm\mu$ is a latent (unobserved) parameter measuring the preference of each author i. $\theta(i) > 0$ indicates a preference for the mostly plus signature
- b_c is the unobserved bias associated with research area c
- If $|\theta| \gg |b|$, individual preferences dominate the need to adapt to a given research area
- "Item-response model": recover invisible traits/factors that may account for observed behaviors.

Sequential and contextual consistency

-  publishes d in category $c_d \in \{\text{phenomenology, theory, …}\}$. What determines which convention she uses? Our assumption:

$P(\sigma_d = +1 | a_d =$  $, c_d) = f(\theta($  $) + b(c_d))$

Author's preference — Effect of research area c_d



- $\theta(i) = \pm\mu$ is a latent (unobserved) parameter measuring the preference of each author i. $\theta(i) > 0$ indicates a preference for the mostly plus signature
- b_c is the unobserved bias associated with research area c
- If $|\theta| \gg |b|$, individual preferences dominate the need to adapt to a given research area
- "Item-response model": recover invisible traits/factors that may account for observed behaviors.
- **Given physicists' choices in their solo-authored papers, we can infer back θ and b using Bayesian inference.**

Sequential and contextual consistency

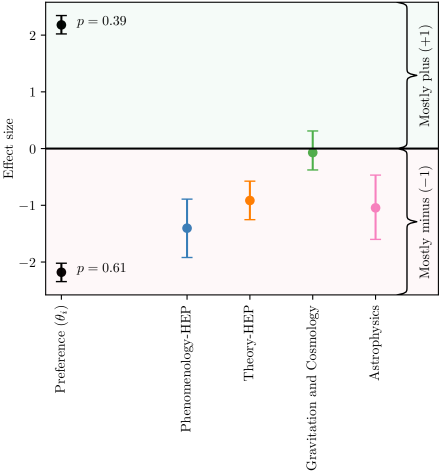


Figure: Sequential consistency (preferences) matter the most, but adaptation to the context also occurs.

Sequential and contextual consistency

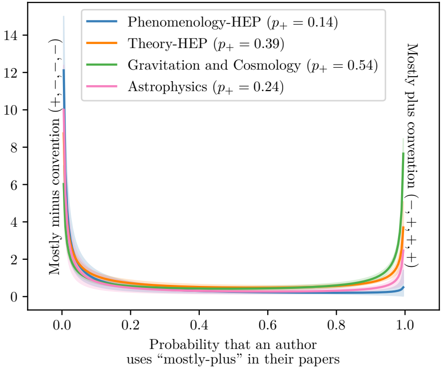


Figure: Physicists tend to always be using the same convention

1. Inverse problems for philosophers and agent-based modelers

2. A case-study of conventions: the metric signature in particle physics
   - How do physicists choose which convention to use in their own papers?
   - How do scientists resolve conflicting preferences in collaborations?
   - How do physicists' preferences get formed?

Inferring preference-aggregation mechanisms in conflicts

How scientists resolve conflicting preferences in collaborations?

Inferring preference-aggregation mechanisms in conflicts

How scientists resolve conflicting preferences in collaborations?

- Focusing on co-authored papers for which:

Inferring preference-aggregation mechanisms in conflicts

How scientists resolve conflicting preferences in collaborations?

- Focusing on co-authored papers for which:
  (i) The metric signature $S_d \in \{-1, +1\}$ of the paper is observed

Inferring preference-aggregation mechanisms in conflicts

How scientists resolve conflicting preferences in collaborations?

- Focusing on co-authored papers for which:
  - (i) The metric signature $S_d \in \{-1, +1\}$ of the paper is observed
  - (ii) The preference of each author $(\sigma_1, \ldots, \sigma_n) \in \{\pm 1\}^n$ is known independently from at least one solo-authored publication

Inferring preference-aggregation mechanisms in conflicts

How scientists resolve conflicting preferences in collaborations?

- Focusing on co-authored papers for which:
  (i) The metric signature $S_d \in \{-1, +1\}$ of the paper is observed
  (ii) The preference of each author $(\sigma_1, \ldots, \sigma_n) \in \{\pm 1\}^n$ is known independently from at least one solo-authored publication
- We can assume different preference aggregation strategies (A_k):

Inferring preference-aggregation mechanisms in conflicts

How scientists resolve conflicting preferences in collaborations?

- Focusing on co-authored papers for which:
  - (i) The metric signature $S_d \in \{-1, +1\}$ of the paper is observed
  - (ii) The preference of each author $(\sigma_1, \ldots, \sigma_n) \in \{\pm 1\}^n$ is known independently from at least one solo-authored publication
- We can assume different preference aggregation strategies (A_k):
  - Dictatorial strategies (the first author, the last author, or another author decides)

Inferring preference-aggregation mechanisms in conflicts

How scientists resolve conflicting preferences in collaborations?

- Focusing on co-authored papers for which:
  - (i) The metric signature $S_d \in \{-1, +1\}$ of the paper is observed
  - (ii) The preference of each author $(\sigma_1, \ldots, \sigma_n) \in \{\pm 1\}^n$ is known independently from at least one solo-authored publication
- We can assume different preference aggregation strategies (A_k):
  - Dictatorial strategies (the first author, the last author, or another author decides)
  - Majoritarian strategy (the majority preference prevails)

Inferring preference-aggregation mechanisms in conflicts

How scientists resolve conflicting preferences in collaborations?

- Focusing on co-authored papers for which:
  (i) The metric signature $S_d \in \{-1, +1\}$ of the paper is observed
  (ii) The preference of each author $(\sigma_1, \ldots, \sigma_n) \in \{\pm 1\}^n$ is known independently from at least one solo-authored publication
- We can assume different preference aggregation strategies (A_k):
  - Dictatorial strategies (the first author, the last author, or another author decides)
  - Majoritarian strategy (the majority preference prevails)
  - Conventional strategy (the signature most common in the target research area prevails)

Inferring preference-aggregation mechanisms in conflicts

How scientists resolve conflicting preferences in collaborations?

- Focusing on co-authored papers for which:
  - (i) The metric signature $S_d \in \{-1, +1\}$ of the paper is observed
  - (ii) The preference of each author $(\sigma_1, \dots, \sigma_n) \in \{\pm 1\}^n$ is known independently from at least one solo-authored publication
- We can assume different preference aggregation strategies (A_k):
  - Dictatorial strategies (the first author, the last author, or another author decides)
  - Majoritarian strategy (the majority preference prevails)
  - Conventional strategy (the signature most common in the target research area prevails)
  - Random/coin-flip (both individual preferences and context are ignored)

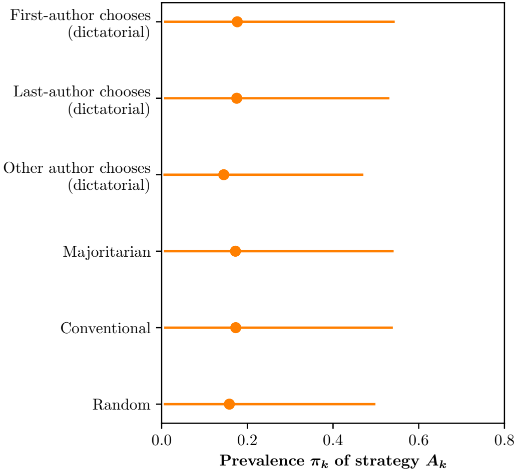
Inferring preference-aggregation mechanisms in conflicts

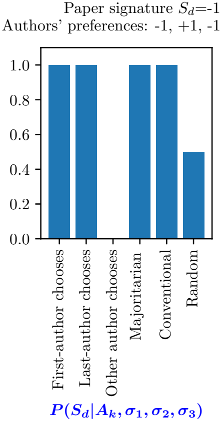
How scientists resolve conflicting preferences in collaborations?

- Focusing on co-authored papers for which:
  (i) The metric signature $S_d \in \{-1, +1\}$ of the paper is observed
  (ii) The preference of each author $(\sigma_1, \ldots, \sigma_n) \in \{\pm 1\}^n$ is known independently from at least one solo-authored publication
- We can assume different preference aggregation strategies (A_k):
  - Dictatorial strategies (the first author, the last author, or another author decides)
  - Majoritarian strategy (the majority preference prevails)
  - Conventional strategy (the signature most common in the target research area prevails)
  - Random/coin-flip (both individual preferences and context are ignored)
- We can estimate the prevalence of each strategy (π_k) given that they predict different outcomes (different probabilities $P(S_d|\sigma_1, \ldots, \sigma_n, A_k)$)

Inferring preference-aggregation mechanisms

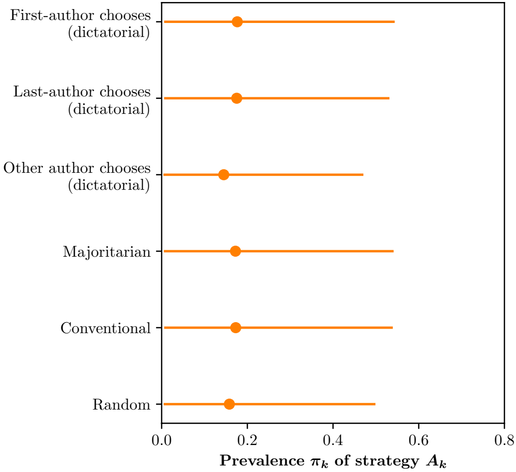
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.

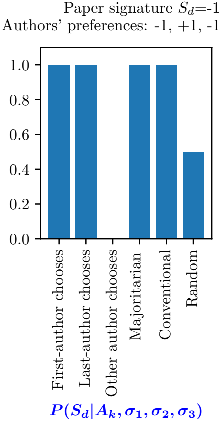




Inferring preference-aggregation mechanisms

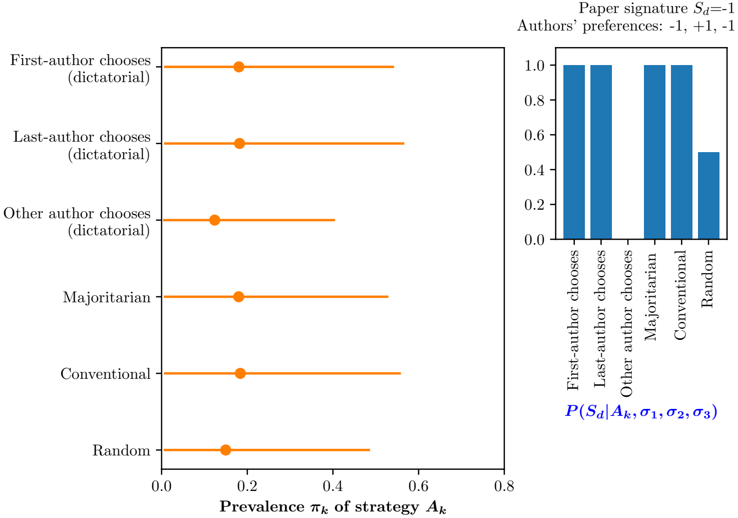
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.





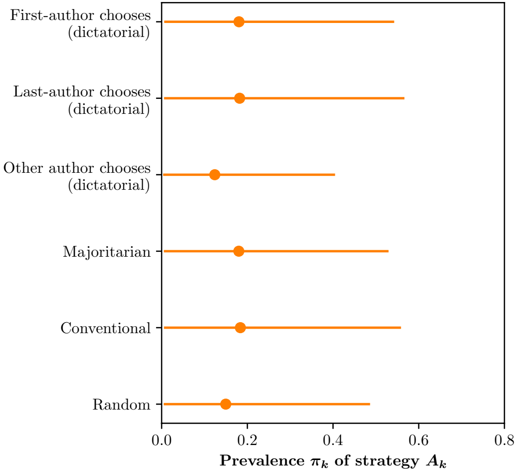
Inferring preference-aggregation mechanisms

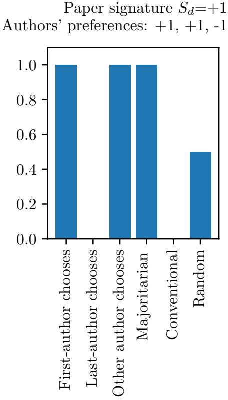
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.



Inferring preference-aggregation mechanisms

Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.

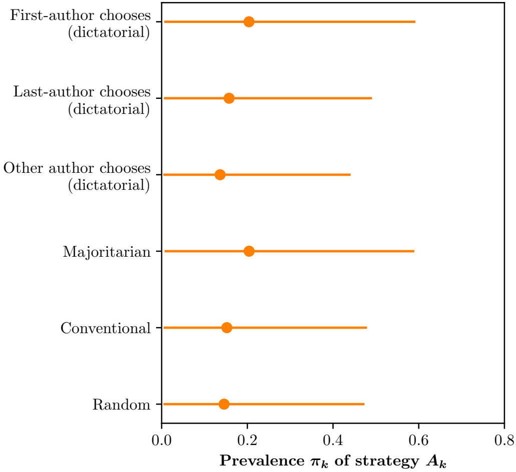


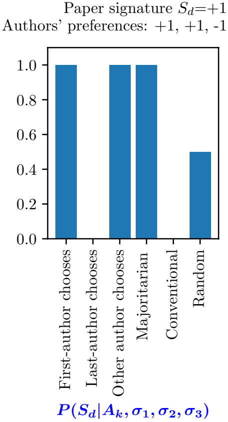


$P(S_d|A_k, \sigma_1, \sigma_2, \sigma_3)$

Inferring preference-aggregation mechanisms

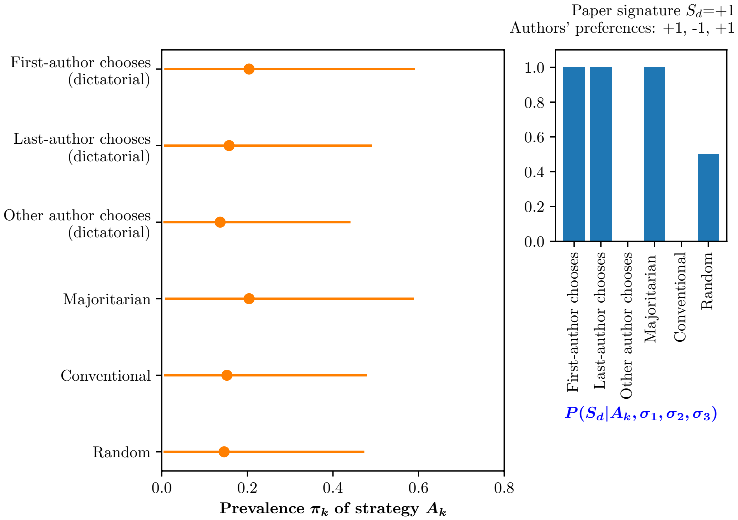
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.





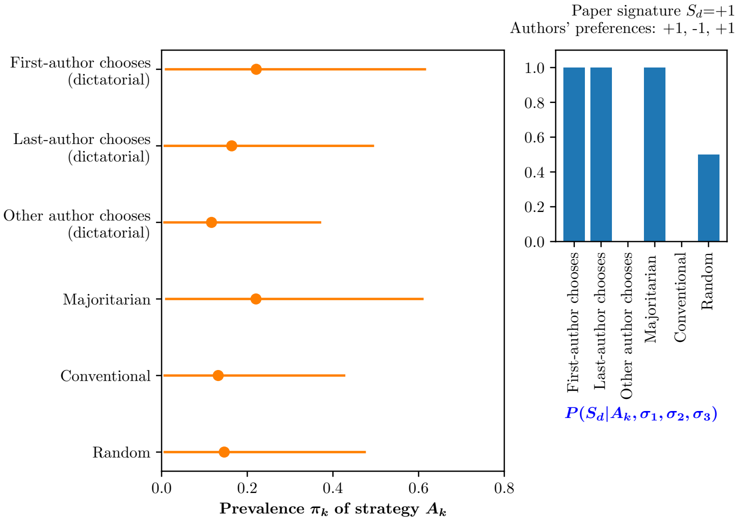
Inferring preference-aggregation mechanisms

Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.



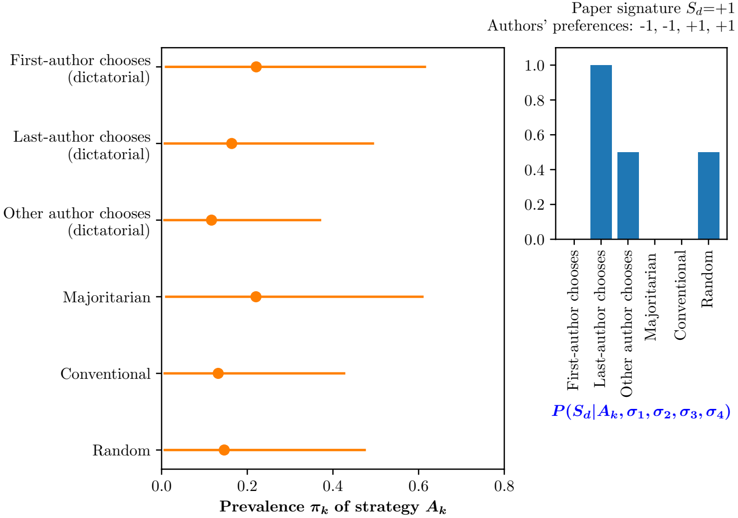
Inferring preference-aggregation mechanisms

Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.



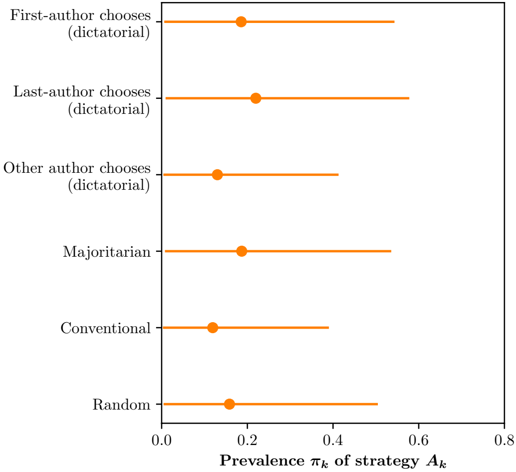
Inferring preference-aggregation mechanisms

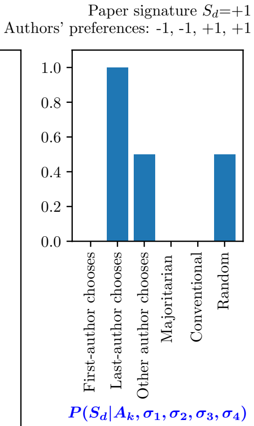
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.



Inferring preference-aggregation mechanisms

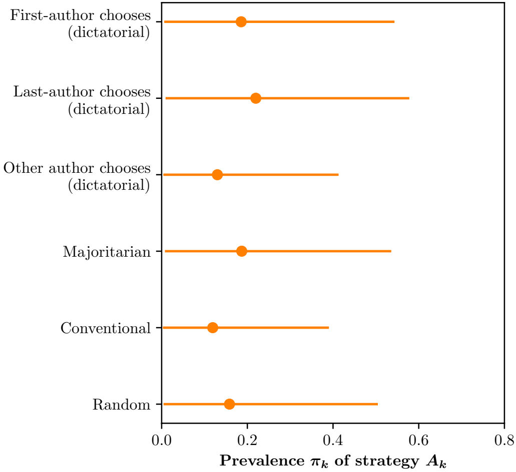
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.

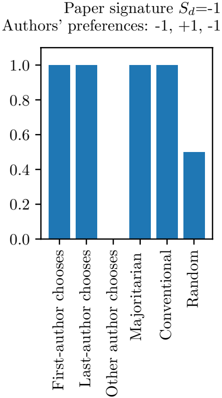




Inferring preference-aggregation mechanisms

Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.

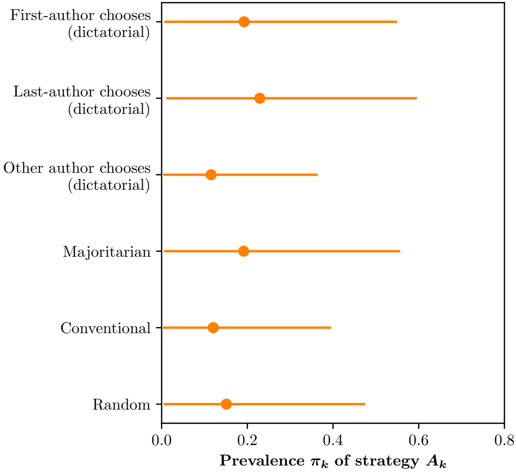


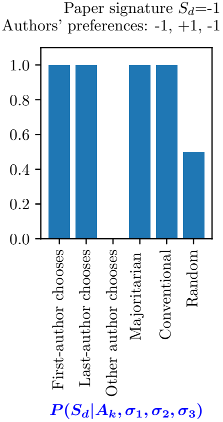


$P(S_d|A_k, \sigma_1, \sigma_2, \sigma_3)$

Inferring preference-aggregation mechanisms

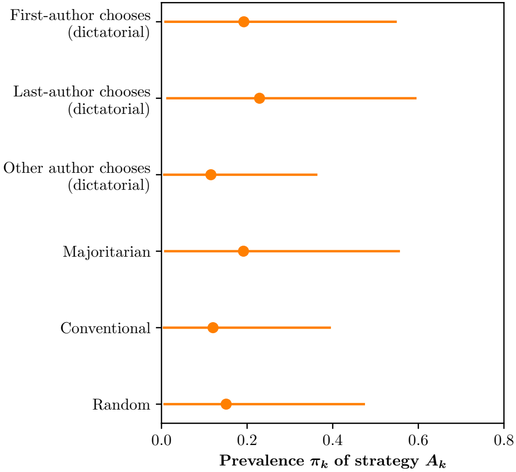
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.

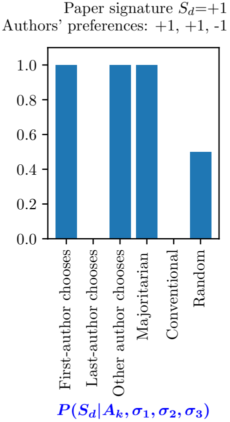




Inferring preference-aggregation mechanisms

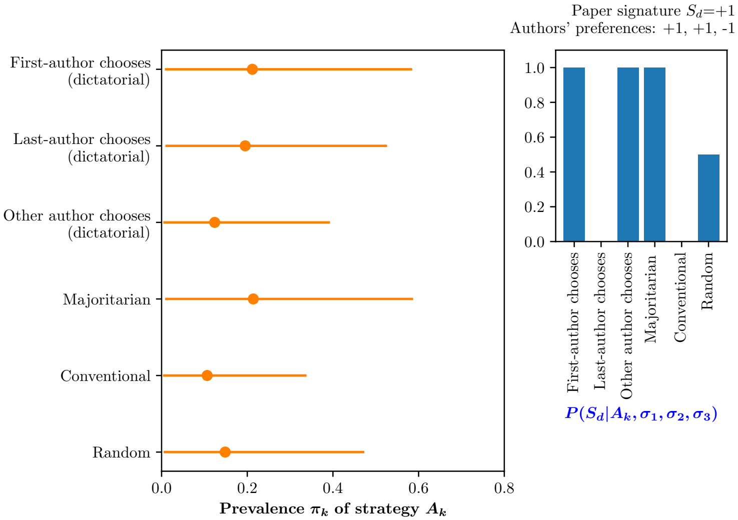
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.





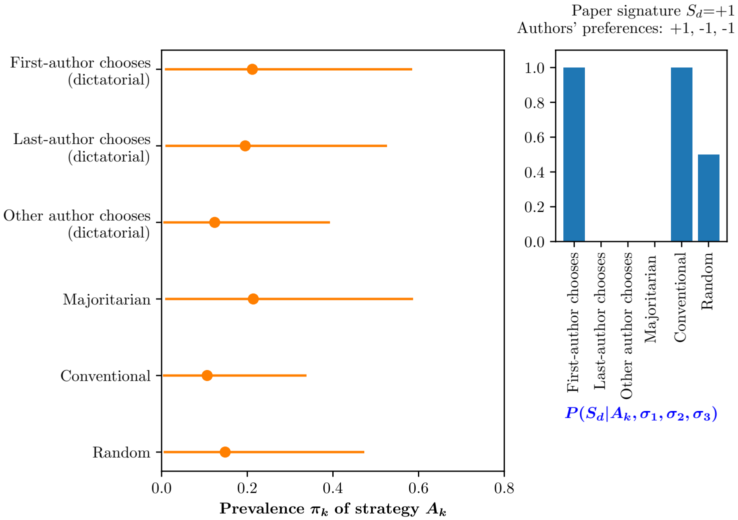
Inferring preference-aggregation mechanisms

Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.



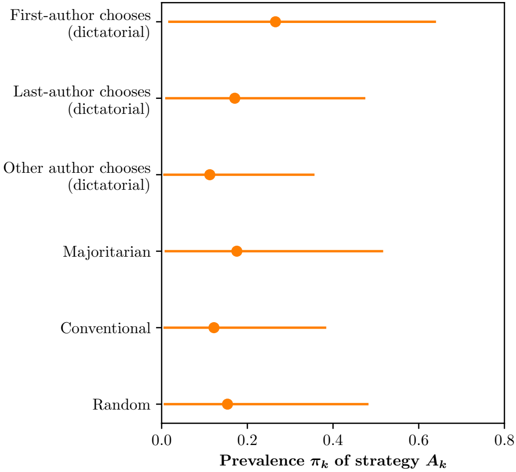
Inferring preference-aggregation mechanisms

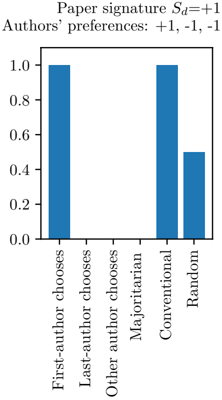
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.



Inferring preference-aggregation mechanisms

Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.

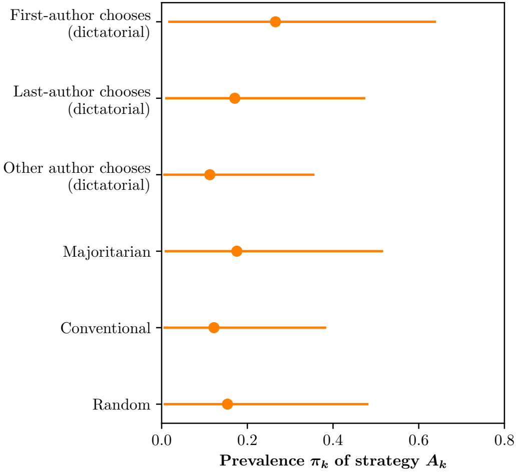


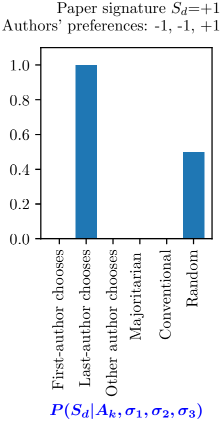


$P(S_d|A_k, \sigma_1, \sigma_2, \sigma_3)$

Inferring preference-aggregation mechanisms

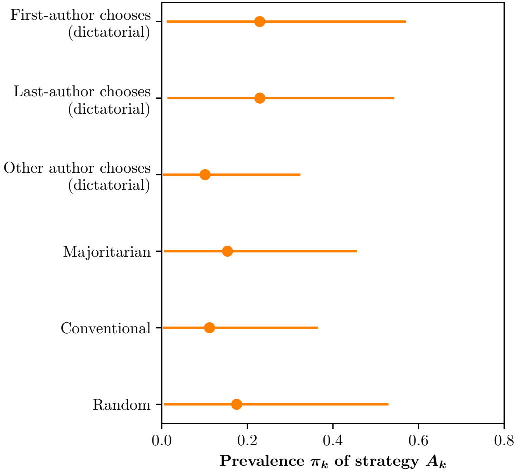
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.

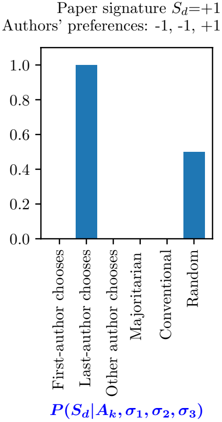




Inferring preference-aggregation mechanisms

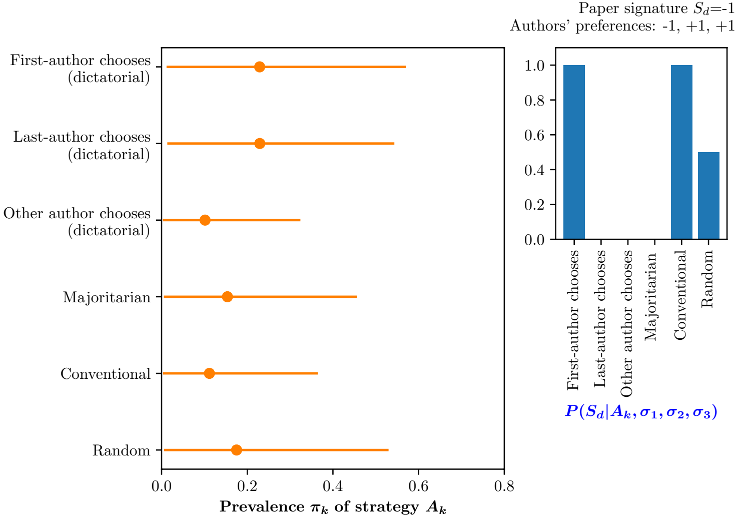
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.





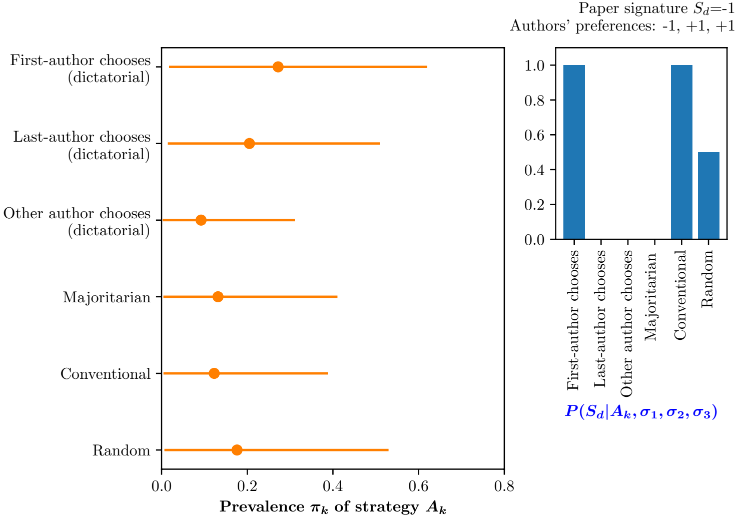
Inferring preference-aggregation mechanisms

Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.



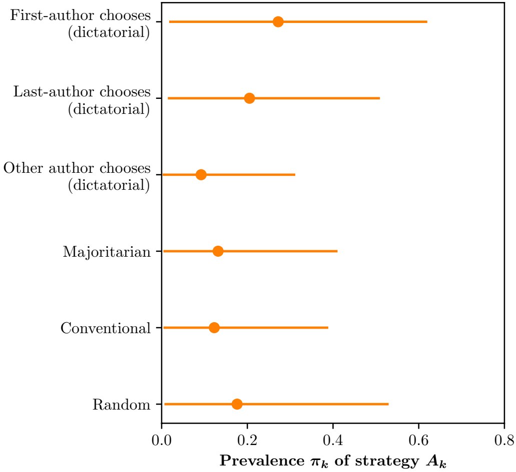
Inferring preference-aggregation mechanisms

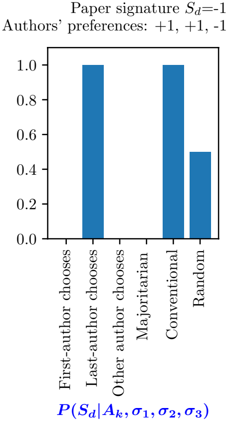
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.



Inferring preference-aggregation mechanisms

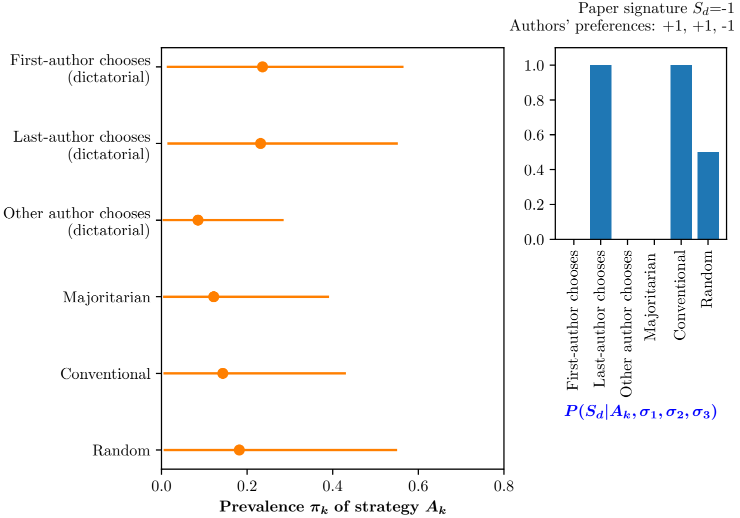
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.





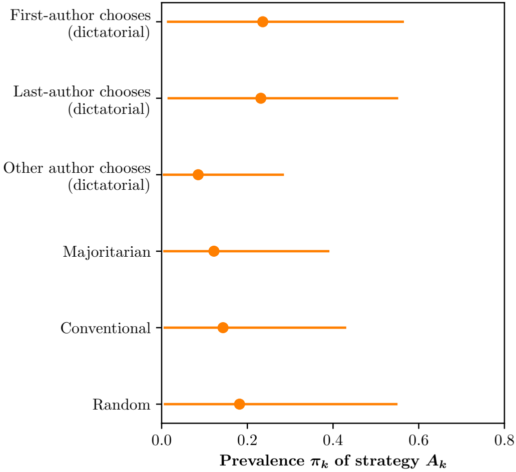
Inferring preference-aggregation mechanisms

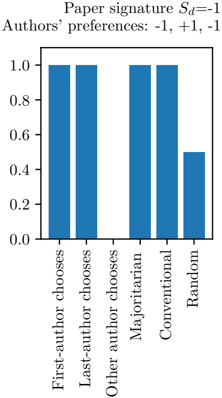
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.



Inferring preference-aggregation mechanisms

Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.

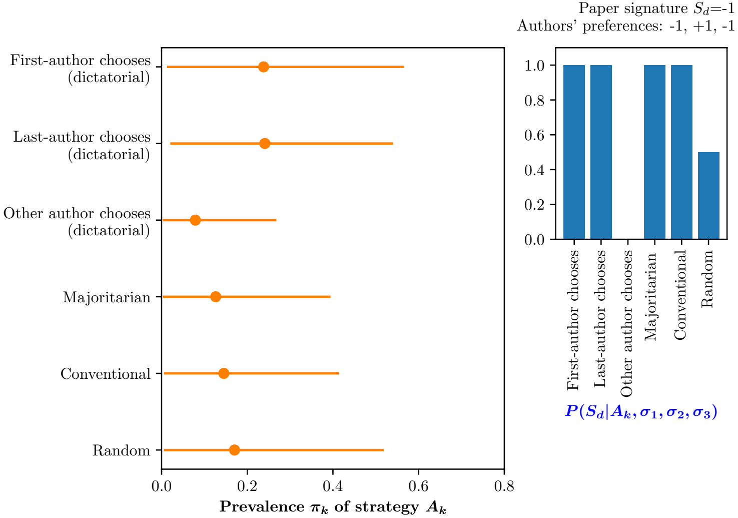




$P(S_d|A_k, \sigma_1, \sigma_2, \sigma_3)$

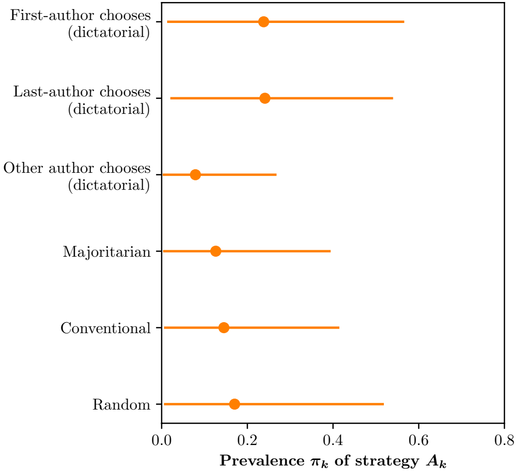
Inferring preference-aggregation mechanisms

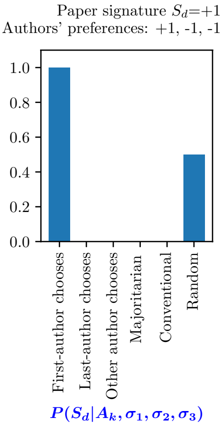
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.



Inferring preference-aggregation mechanisms

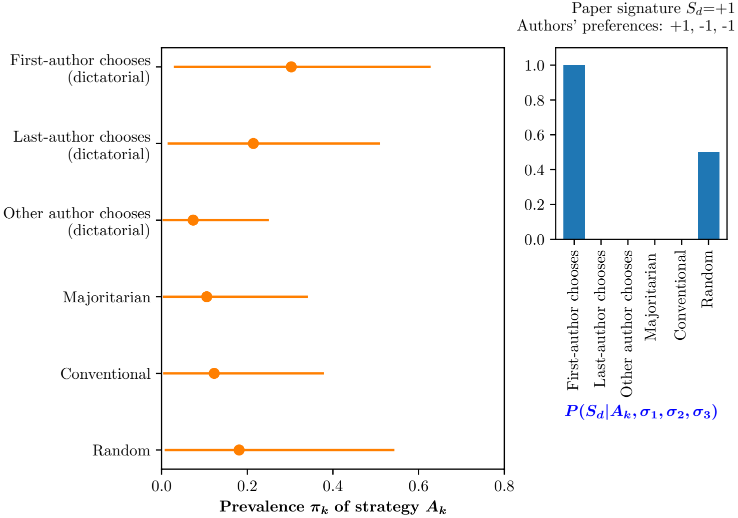
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.





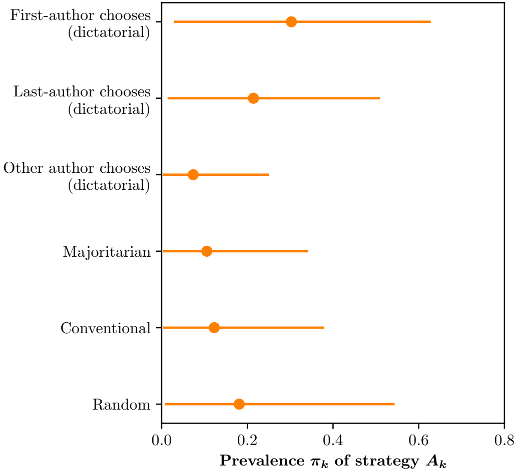
Inferring preference-aggregation mechanisms

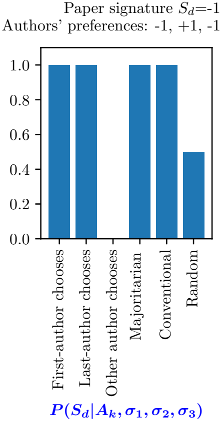
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.



Inferring preference-aggregation mechanisms

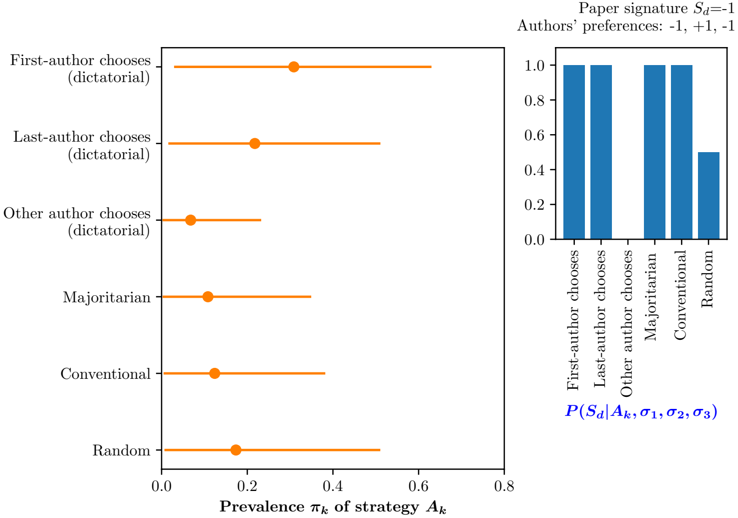
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.





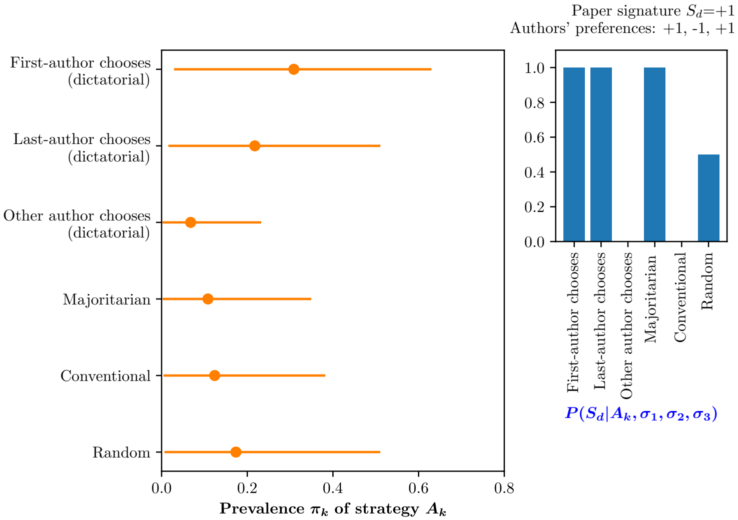
Inferring preference-aggregation mechanisms

Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.



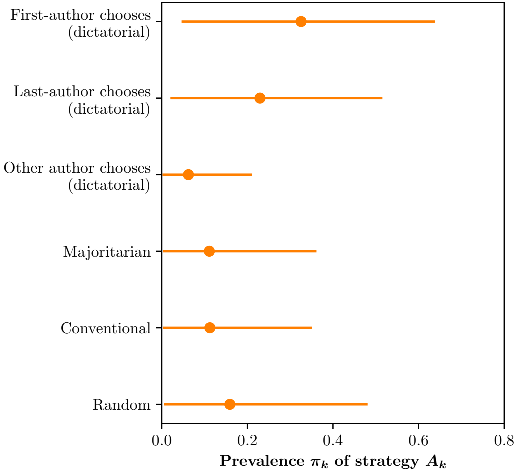
Inferring preference-aggregation mechanisms

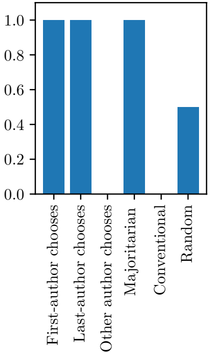
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.



Inferring preference-aggregation mechanisms

Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.

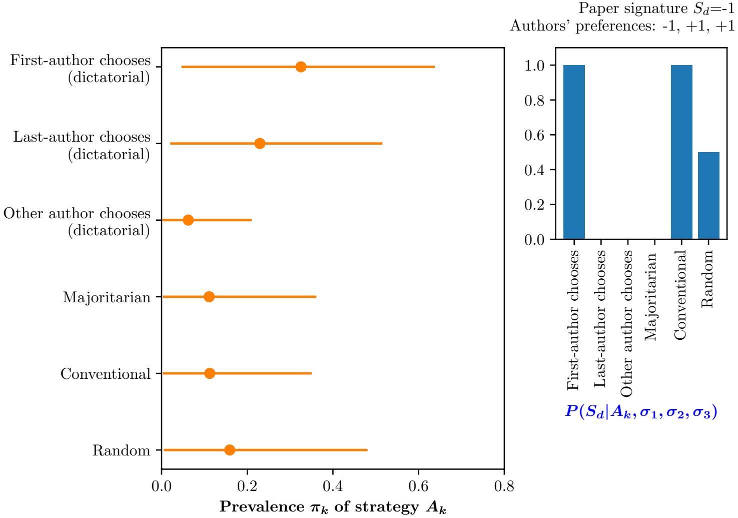




$P(S_d | A_k, \sigma_1, \sigma_2, \sigma_3)$

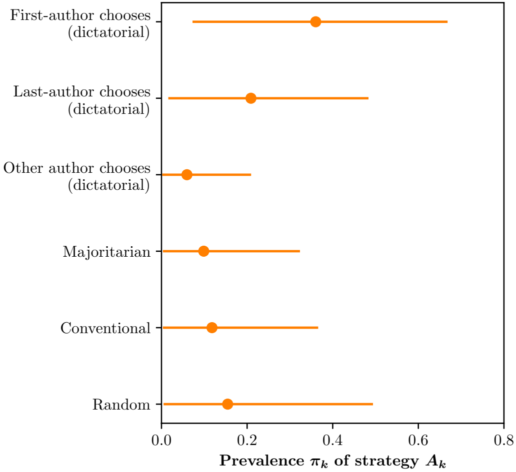
Inferring preference-aggregation mechanisms

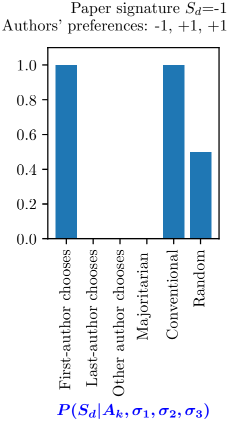
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.



Inferring preference-aggregation mechanisms

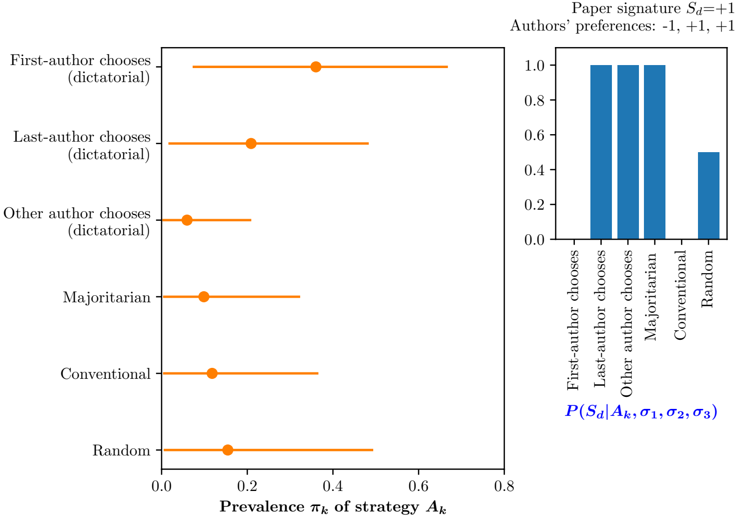
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.





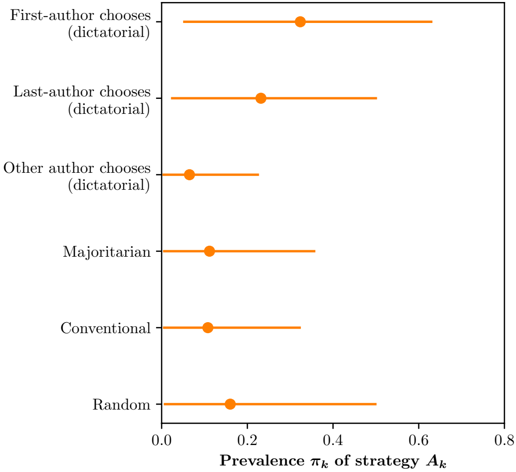
Inferring preference-aggregation mechanisms

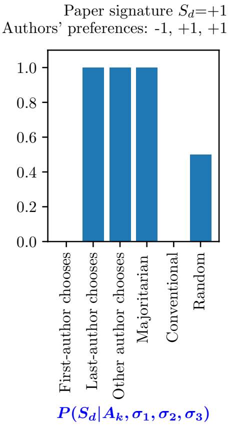
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.



Inferring preference-aggregation mechanisms

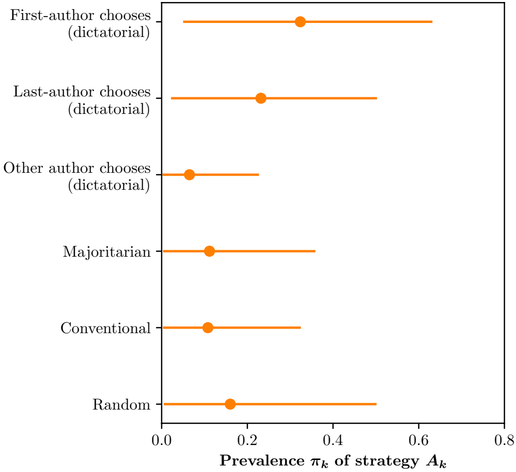
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.

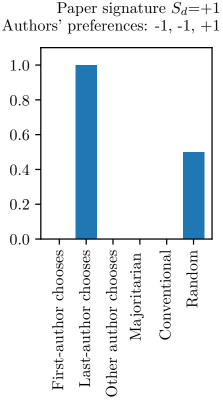




Inferring preference-aggregation mechanisms

Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.

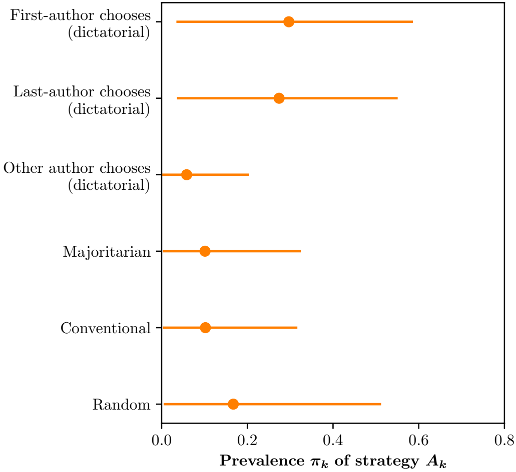


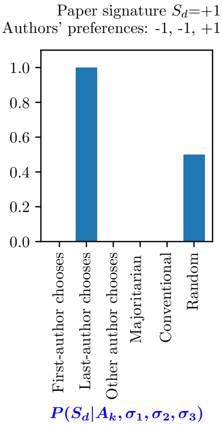


$P(S_d|A_k, \sigma_1, \sigma_2, \sigma_3)$

Inferring preference-aggregation mechanisms

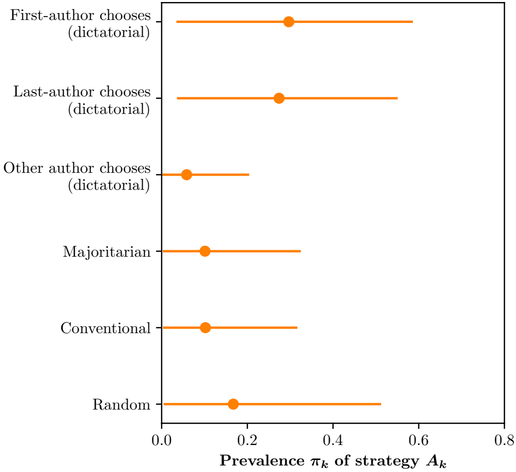
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.

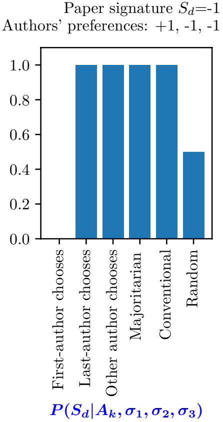




Inferring preference-aggregation mechanisms

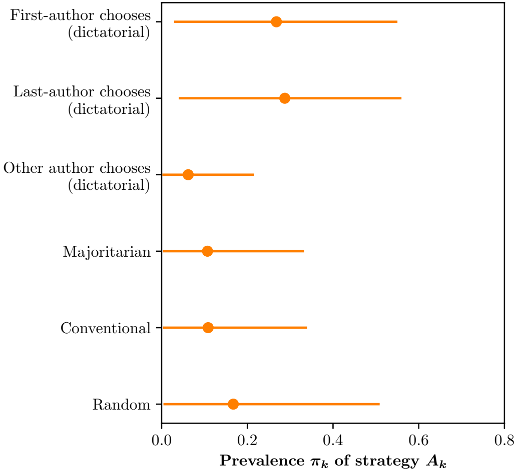
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.

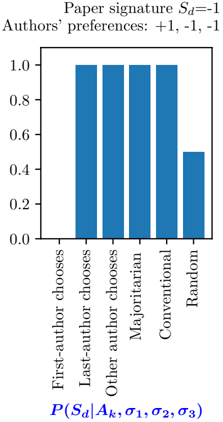




Inferring preference-aggregation mechanisms

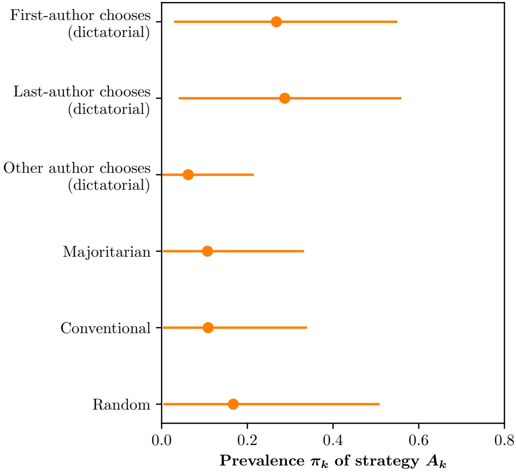
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.

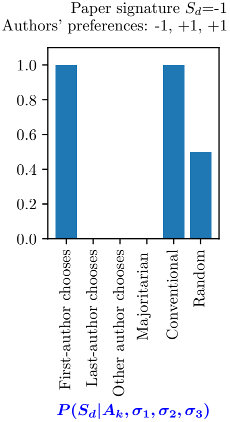




Inferring preference-aggregation mechanisms

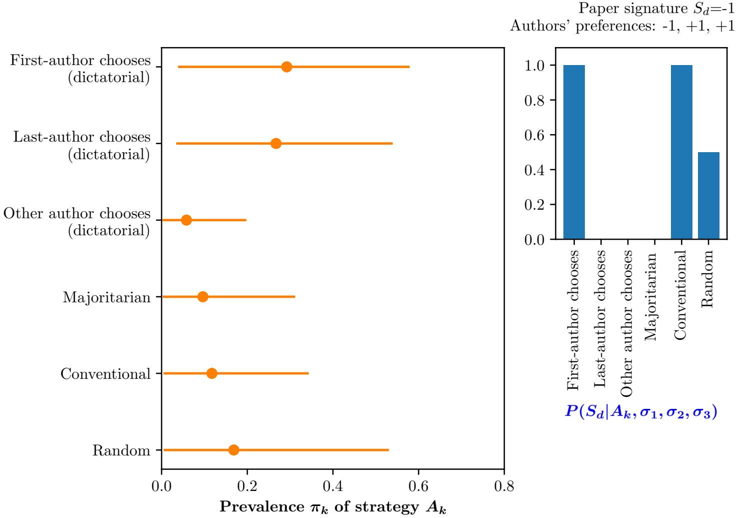
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.





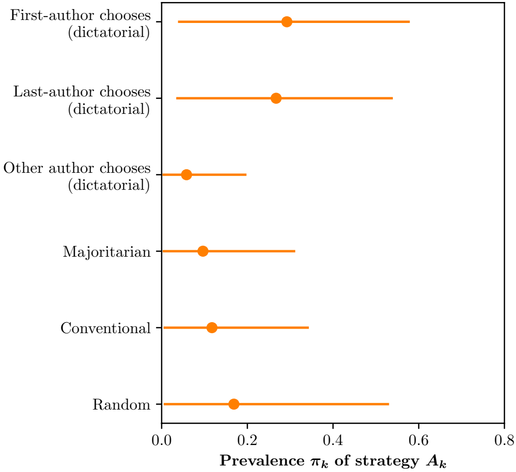
Inferring preference-aggregation mechanisms

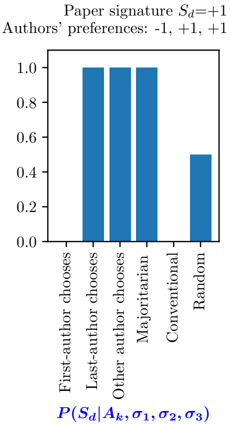
Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.



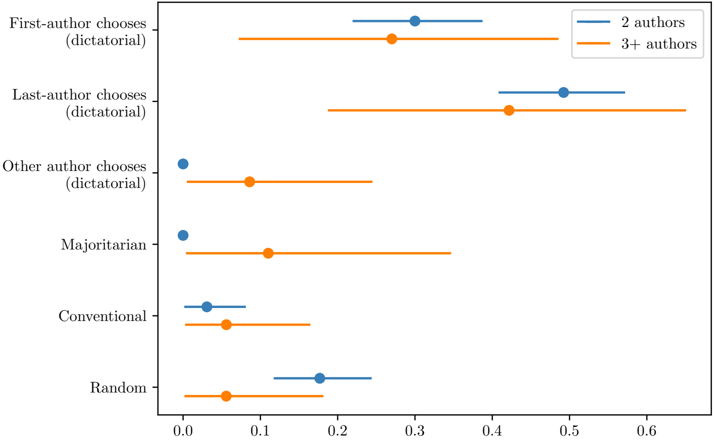
Inferring preference-aggregation mechanisms

Each paper brings a bit more information about π_k, the prevalence of an aggregation strategy A_k.





Prevalence of each preference-aggregation strategy



1. Inverse problems for philosophers and agent-based modelers

2. A case-study of conventions: the metric signature in particle physics
    - How do physicists choose which convention to use in their own papers?
    - How do scientists resolve conflicting preferences in collaborations?
    - How do physicists' preferences get formed?

Authors' preferences

Observed outcome: the preference of each author,

$$O_{\text{obs}} = (\sigma_1, \ldots, \sigma_n), \sigma \in \{-1, +1\}$$

($n = 2\,277$ authors)

How do physicists' preferences get formed?

- Let's assume three models of the formation of physicists' preference towards the convention:
  1. A **"strategic agent" model** (M_1) assuming that individuals navigate three costs (coordination costs, inconsistency costs, and maladaptation costs) depending on their collaborators' preferences and the research areas in which they publish.
  2. A **global cultural transmission model** (M_2), in which physicists settle once and for all for a specific convention with a certain probability that depends on their primary research area (textbooks?)
  3. A **local cultural transmission model** (M_3), in which physicists copy the preference of their first collaborator.
- Which of these is more plausible given the observed patterns of preferences?

Example: the strategic agent model (M_1)

The model M_1 has multiple unknown parameters:

- c_s: the cost of switching from one convention to another
- c_c: the cost of disagreeing with co-authors
- c_r the cost of using a suboptimal convention in a given research area

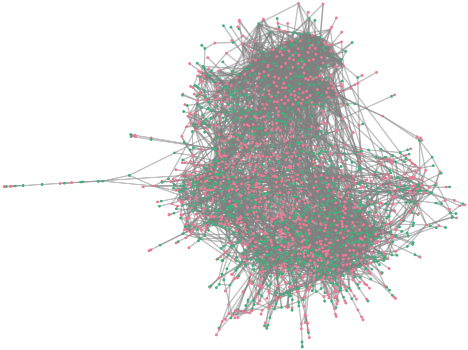
The **outcome** O_{sim} is the joint value of each author's preference:
$O_{\mathrm{sim}} = (\sigma_1, \ldots, \sigma_n)$ where $\sigma_i = \pm 1$

Example: the strategic agent model (M_1)

The model M_1 has multiple unknown parameters:

- c_s: the cost of switching from one convention to another
- c_c: the cost of disagreeing with co-authors
- c_r the cost of using a suboptimal convention in a given research area

The **outcome** O_{sim} is the joint value of each author's preference:
$O_{\text{sim}} = (\sigma_1, \ldots, \sigma_n)$ where $\sigma_i = \pm 1$



Simulation-based inference

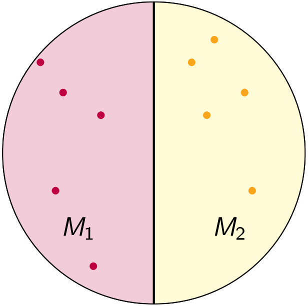
$$P(M_1|O) = \overbrace{P(O|M_1)}\frac{P(M_1)}{P(O)} \tag{4}$$

Simulation-based inference

$$P(M_1|O) = \overbrace{P(O|M_1)}^{\substack{\text{Unknown} \\ \text{in ABMs!}}} \frac{P(M_1)}{P(O)} \tag{4}$$

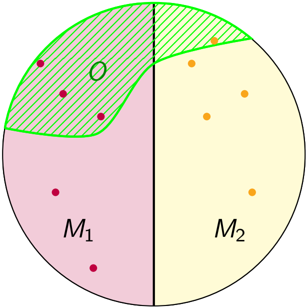
Simulation-based inference

$$\color{red}{P(M_1|O)} = \color{blue}{\overbrace{P(O|M_1)}^{\substack{\text{Unknown} \\ \text{in ABMs!}}}} \frac{P(M_1)}{P(O)} \tag{4}$$



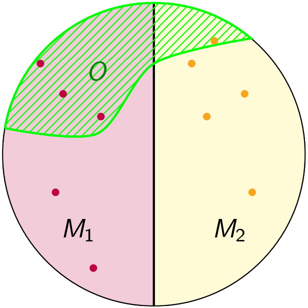
Simulation-based inference

$$P(M_1|O) = \overbrace{P(O|M_1)}^{\text{Unknown in ABMs!}} \frac{P(M_1)}{P(O)} \tag{4}$$



Simulation-based inference

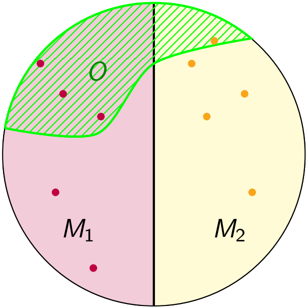
$$\color{red}{P(M_1|O)} = \overbrace{\color{blue}{P(O|M_1)}}^{\text{Unknown in ABMs!}} \frac{P(M_1)}{P(O)} \tag{4}$$



$$\color{blue}{P(O|M_1)} = \frac{\quad}{\quad} \simeq \frac{3}{5}$$

Simulation-based inference

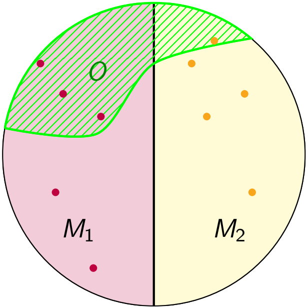
$$P(M_1|O) = \overbrace{P(O|M_1)}^{\text{Unknown in ABMs!}} \frac{P(M_1)}{P(O)} \tag{4}$$



$$P(O|M_1) = \frac{\boxed{}}{\boxed{}} \simeq \frac{3}{5}$$

$$P(O|M_2) = \frac{\boxed{}}{\boxed{}} \simeq \frac{1}{5}$$

$$P(M_1|O) = \overbrace{P(O|M_1)}^{\text{Unknown in ABMs!}} \frac{P(M_1)}{P(O)} \tag{4}$$



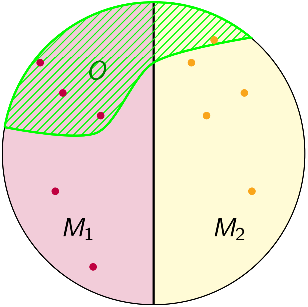
$$P(O|M_1) = \simeq \frac{3}{5}$$

$$P(O|M_2) = \simeq \frac{1}{5}$$

$$P(M_1|O) = P(O|M_1)\frac{P(M_1)}{P(O)} = =$$

Simulation-based inference

$$P(M_1|O) = \overbrace{P(O|M_1)}^{\text{Unknown in ABMs!}} \frac{P(M_1)}{P(O)} \tag{4}$$



$$P(O|M_1) = \simeq \frac{3}{5}$$

$$P(O|M_2) = \simeq \frac{1}{5}$$

$$P(M_1|O) = P(O|M_1)\frac{P(M_1)}{P(O)} = =$$

$$P(M_2|O) = P(O|M_2)\frac{P(M_2)}{P(O)} = =$$

Curse of dimensionality in simulation-based inference

- At least some simulations O_s must match the observed outcome O, which means matching each of the 2 277 authors' preferences at the same time!

Curse of dimensionality in simulation-based inference

- At least some simulations O_s must match the observed outcome O, which means matching each of the 2 277 authors' preferences at the same time!
- Virtually impossible (95% chance of predicting any individual preference σ_i correctly $\Rightarrow 2 \times 10^{-51}$ of getting all of them right at once)

Curse of dimensionality in simulation-based inference

- At least some simulations O_s must match the observed outcome O, which means matching each of the 2 277 authors' preferences at the same time!
- Virtually impossible (95% chance of predicting any individual preference σ_i correctly $\Rightarrow 2 \times 10^{-51}$ of getting all of them right at once)
- The data has *too many dimensions* $\Rightarrow$ **"curse of dimensionality"**

Curse of dimensionality in simulation-based inference

- At least some simulations O_s must match the observed outcome O, which means matching each of the 2 277 authors' preferences at the same time!
- Virtually impossible (95% chance of predicting any individual preference σ_i correctly $\Rightarrow 2 \times 10^{-51}$ of getting all of them right at once)
- The data has *too many dimensions* $\Rightarrow$ **"curse of dimensionality"**
- Solution: "conditioning" on **summary statistics** rather than the entire data.

Curse of dimensionality in simulation-based inference

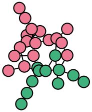
- At least some simulations O_s must match the observed outcome O, which means matching each of the 2 277 authors' preferences at the same time!
- Virtually impossible (95% chance of predicting any individual preference σ_i correctly $\Rightarrow 2 \times 10^{-51}$ of getting all of them right at once)
- The data has *too many dimensions* $\Rightarrow$ **"curse of dimensionality"**
- Solution: "conditioning" on **summary statistics** rather than the entire data.
- Summary statistics are **low-dimensional descriptions of the data** that capture essential features. e.g.:

$$m = \frac{1}{n}\left|\sum_{i=1}^{n} \sigma_i\right| \text{ (where } \sigma_i = \pm 1) \tag{5}$$

Curse of dimensionality in simulation-based inference

- At least some simulations O_s must match the observed outcome O, which means matching each of the 2 277 authors' preferences at the same time!
- Virtually impossible (95% chance of predicting any individual preference σ_i correctly $\Rightarrow 2 \times 10^{-51}$ of getting all of them right at once)
- The data has *too many dimensions* $\Rightarrow$ **"curse of dimensionality"**
- Solution: "conditioning" on **summary statistics** rather than the entire data.
- Summary statistics are **low-dimensional descriptions of the data** that capture essential features. e.g.:

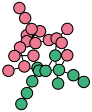
$$m = \frac{1}{n} \left| \sum_{i=1}^{n} \sigma_i \right| \text{ (where } \sigma_i = \pm 1\text{)} \tag{5}$$

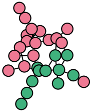
$O_{\text{obs}} =$  $\rightarrow m_{\text{obs}} = 0.03$

Curse of dimensionality in simulation-based inference

- At least some simulations O_s must match the observed outcome O, which means matching each of the 2 277 authors' preferences at the same time!
- Virtually impossible (95% chance of predicting any individual preference σ_i correctly $\Rightarrow 2 \times 10^{-51}$ of getting all of them right at once)
- The data has *too many dimensions* $\Rightarrow$ **"curse of dimensionality"**
- Solution: "conditioning" on **summary statistics** rather than the entire data.
- Summary statistics are **low-dimensional descriptions of the data** that capture essential features. e.g.:

$$m = \frac{1}{n} \left| \sum_{i=1}^{n} \sigma_i \right| \text{ (where } \sigma_i = \pm 1) \tag{5}$$

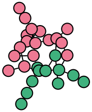
$O_{\text{obs}} =$  $\rightarrow m_{\text{obs}} = 0.03$

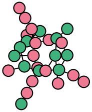
$O_{\text{sim}} =$  $\rightarrow m_{\text{sim}} = 0.03$

Curse of dimensionality in simulation-based inference

- At least some simulations O_s must match the observed outcome O, which means matching each of the 2 277 authors' preferences at the same time!
- Virtually impossible (95% chance of predicting any individual preference σ_i correctly $\Rightarrow 2 \times 10^{-51}$ of getting all of them right at once)
- The data has *too many dimensions* $\Rightarrow$ **"curse of dimensionality"**
- Solution: "conditioning" on **summary statistics** rather than the entire data.
- Summary statistics are **low-dimensional descriptions of the data** that capture essential features. e.g.:

$$m = \frac{1}{n} \left| \sum_{i=1}^{n} \sigma_i \right| \text{ (where } \sigma_i = \pm 1) \tag{5}$$

$O_{\text{obs}} =$  $\rightarrow m_{\text{obs}} = 0.03$

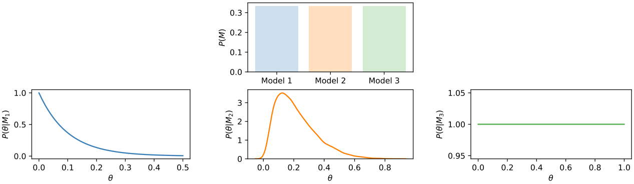
$O_{\text{sim}} =$  $\rightarrow m_{\text{sim}} = 0.03$

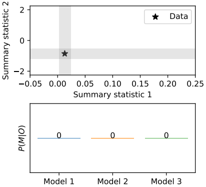
Summary statistics in simulation-based inference

There are two main approaches for choosing adequate summary statistics:

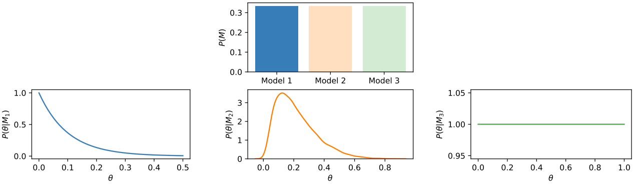
1. Hand-picking interpretable summary statistics based on our own intuitions.
2. Using sophisticated methods to learn statistically optimal (but potentially un-interpretable) summary statistics. Optimal summary statistics reduce our posterior uncertainty given a fixed amount of data.

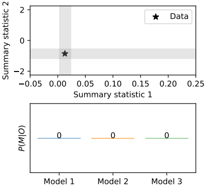
Simulation-based inference with summary statistics



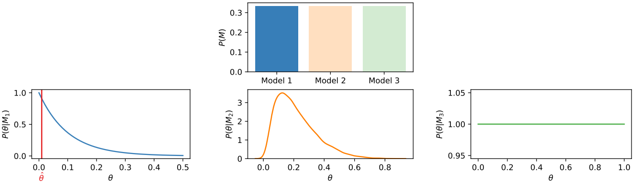


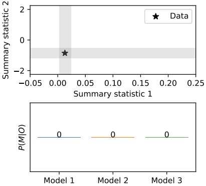
Simulation-based inference with summary statistics



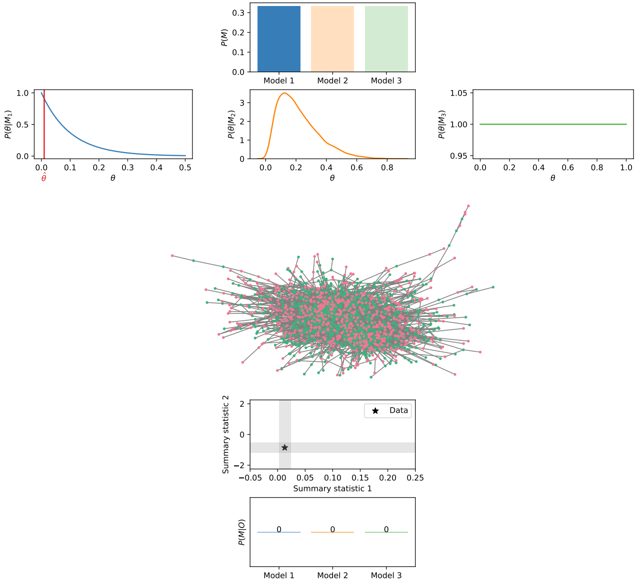


Simulation-based inference with summary statistics

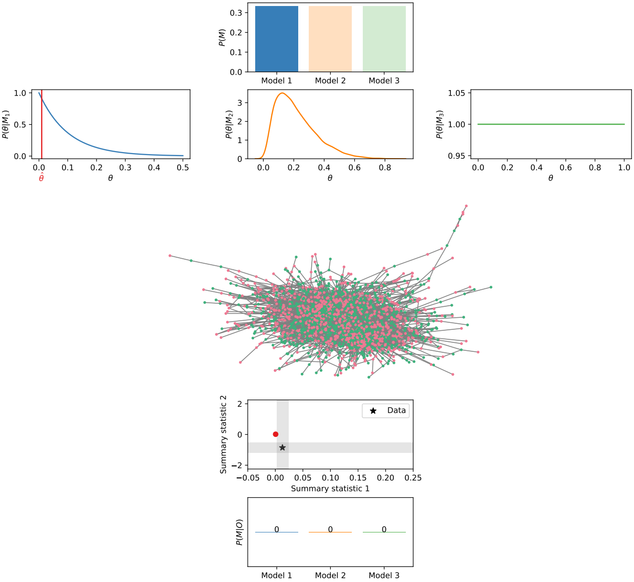




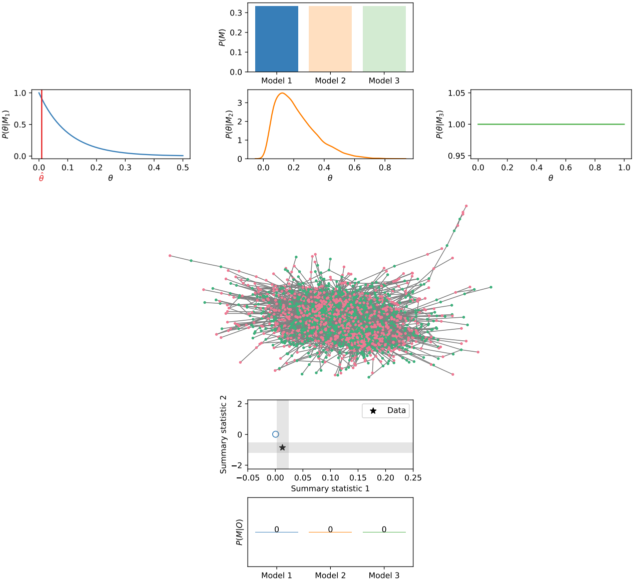
Simulation-based inference with summary statistics



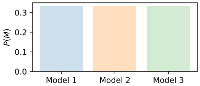
Simulation-based inference with summary statistics

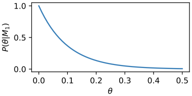


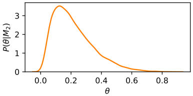
Simulation-based inference with summary statistics

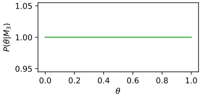


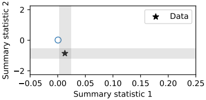
Simulation-based inference with summary statistics

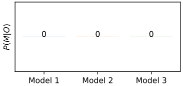




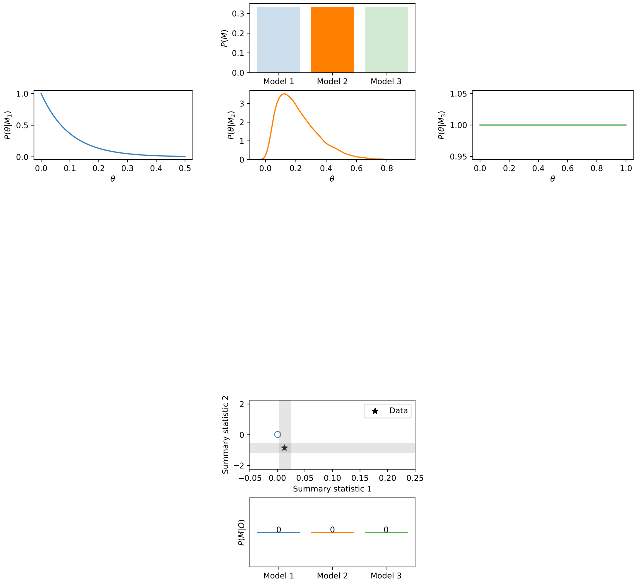




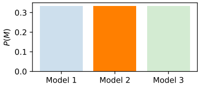


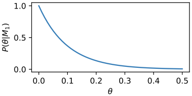


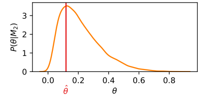
Simulation-based inference with summary statistics

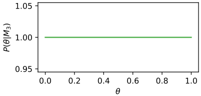


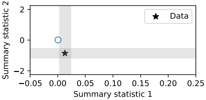
Simulation-based inference with summary statistics

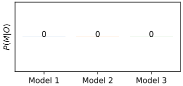




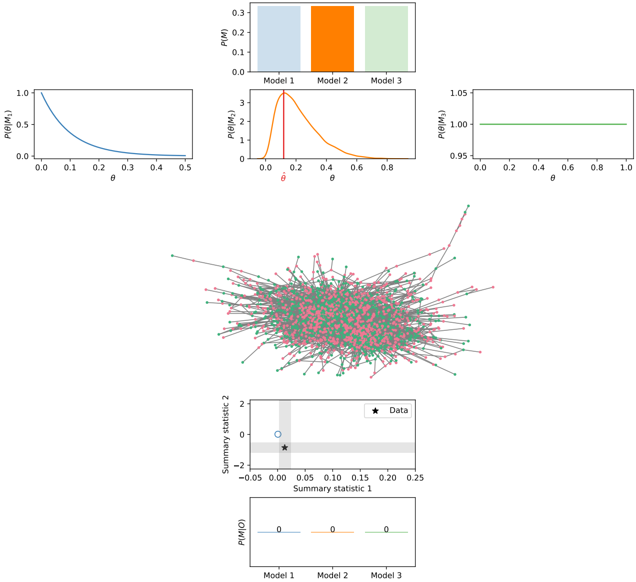




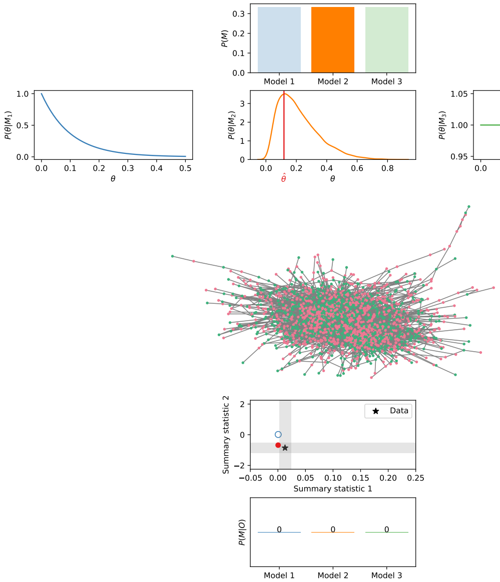




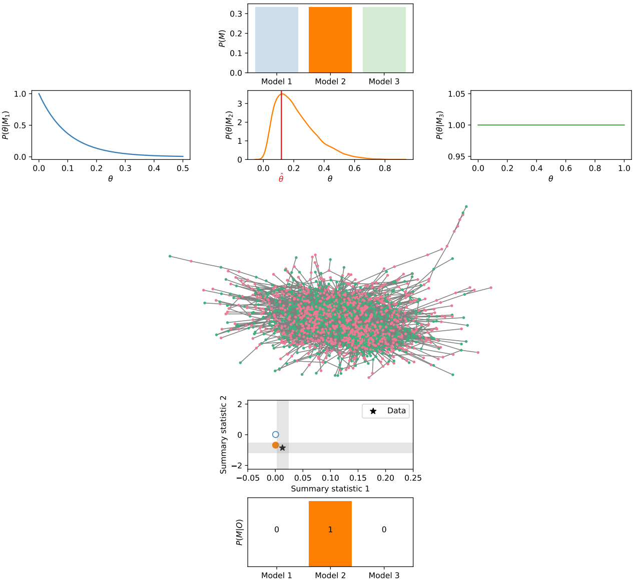
Simulation-based inference with summary statistics



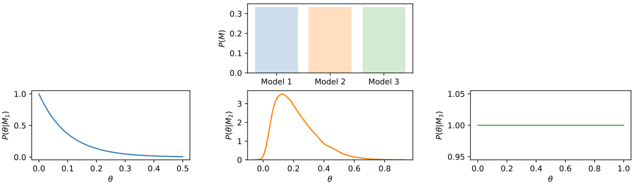
Simulation-based inference with summary statistics

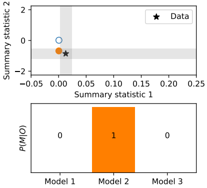


Simulation-based inference with summary statistics

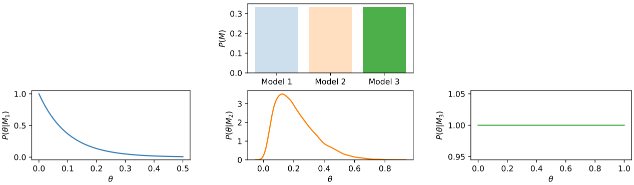


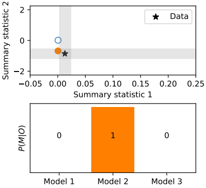
Simulation-based inference with summary statistics



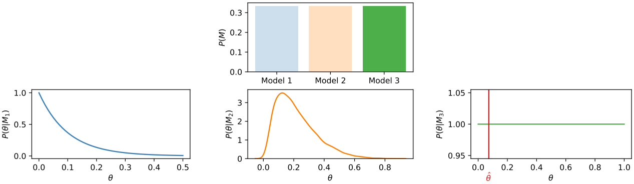


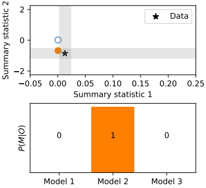
Simulation-based inference with summary statistics



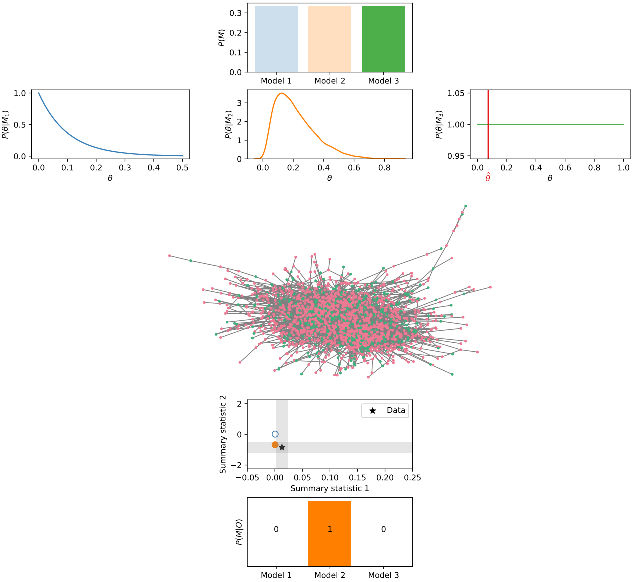


Simulation-based inference with summary statistics

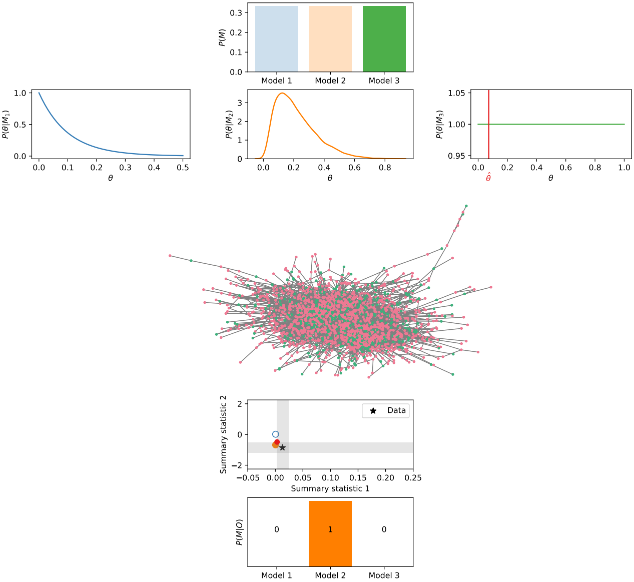




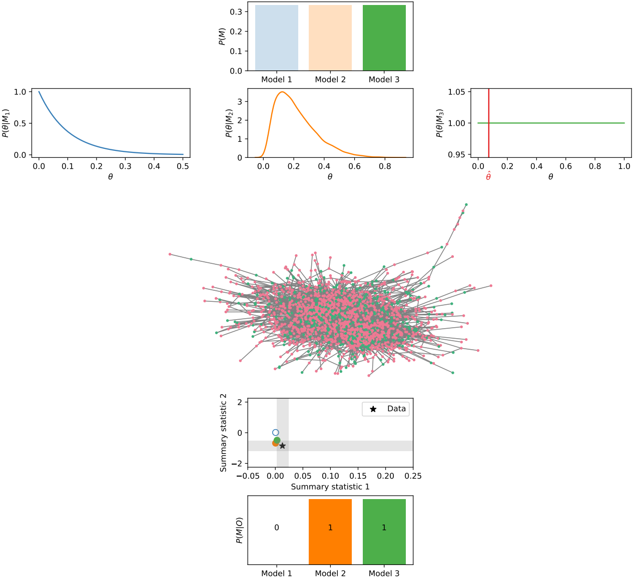
Simulation-based inference with summary statistics



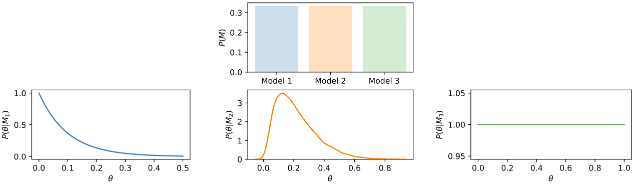
Simulation-based inference with summary statistics

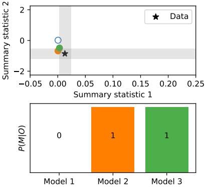


Simulation-based inference with summary statistics

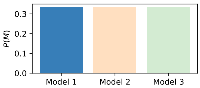


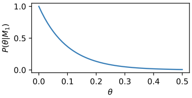
Simulation-based inference with summary statistics

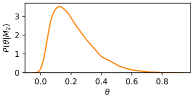


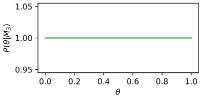


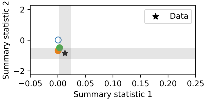
Simulation-based inference with summary statistics

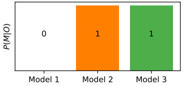




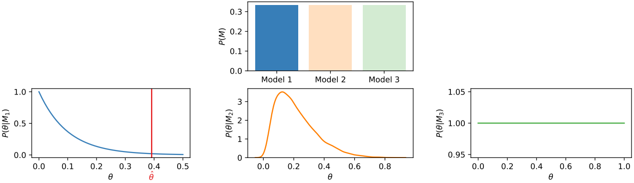


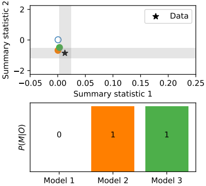




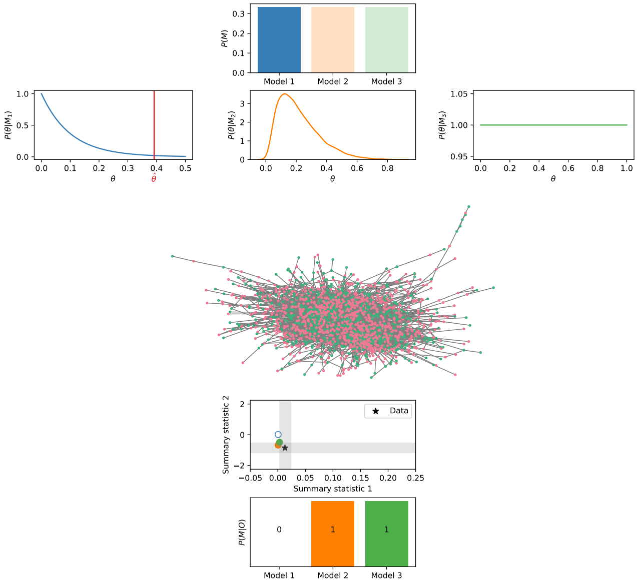


Simulation-based inference with summary statistics

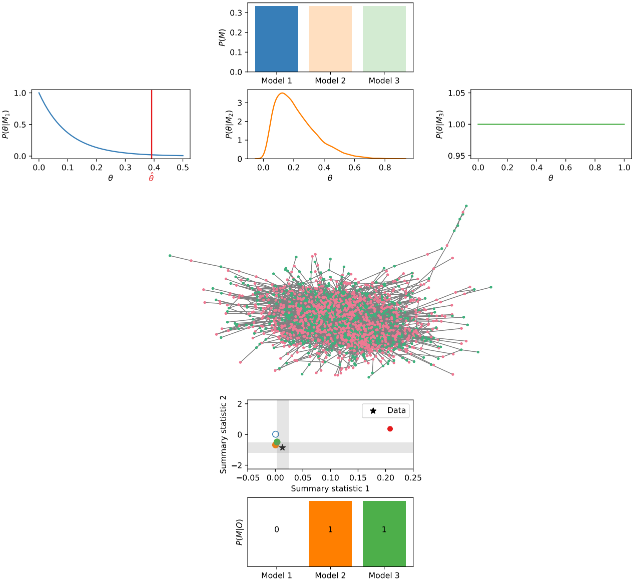




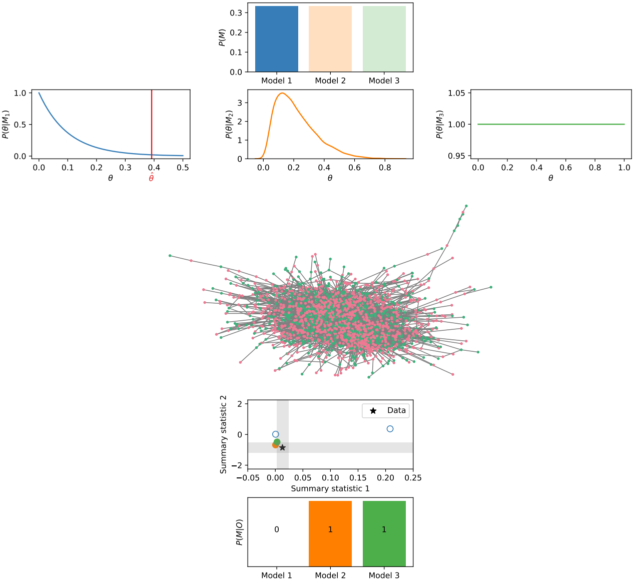
Simulation-based inference with summary statistics



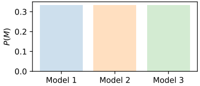
Simulation-based inference with summary statistics

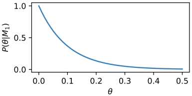


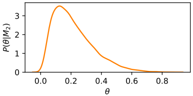
Simulation-based inference with summary statistics

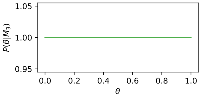


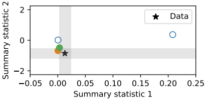
Simulation-based inference with summary statistics

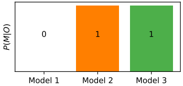




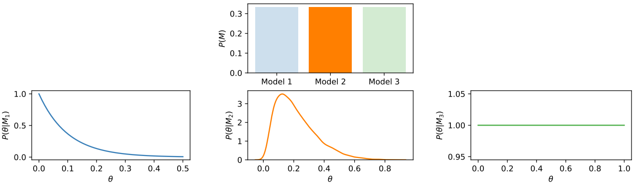


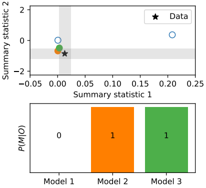




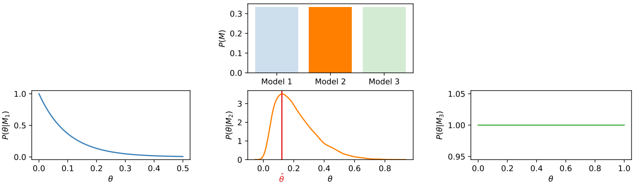


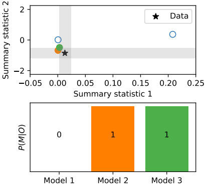
Simulation-based inference with summary statistics



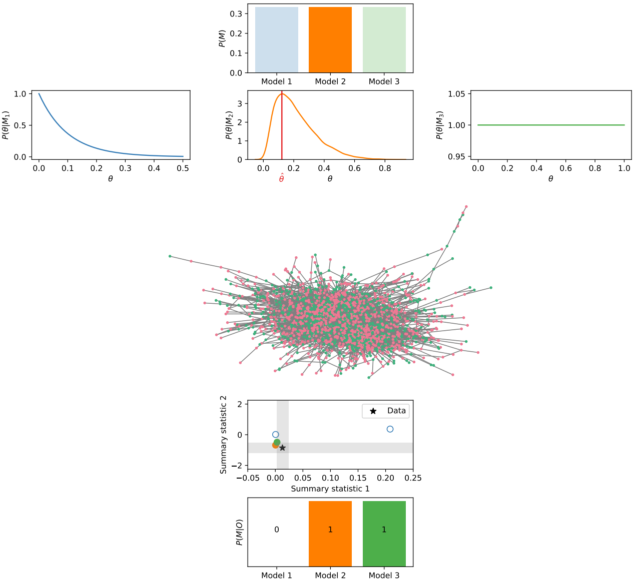


Simulation-based inference with summary statistics

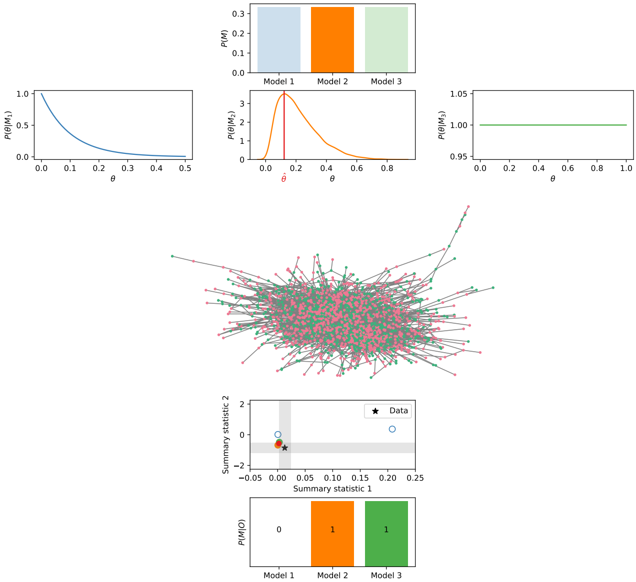




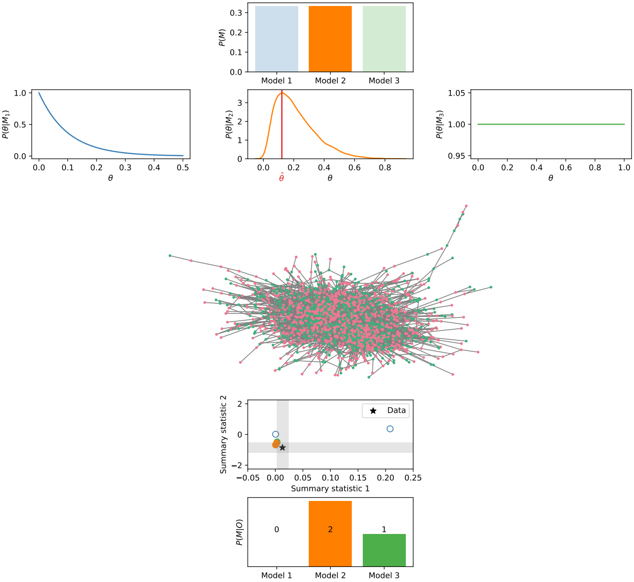
Simulation-based inference with summary statistics



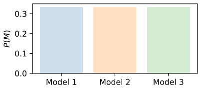
Simulation-based inference with summary statistics

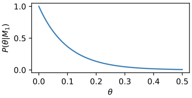


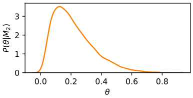
Simulation-based inference with summary statistics

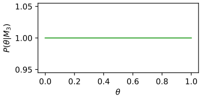


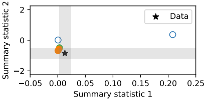
Simulation-based inference with summary statistics

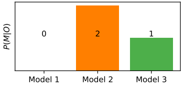




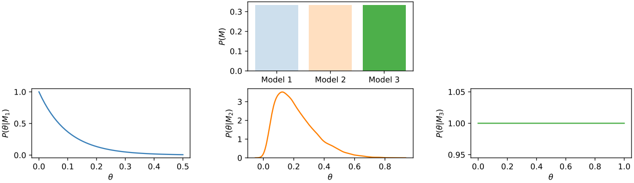


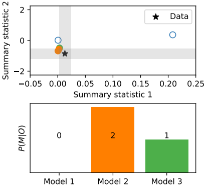




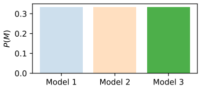


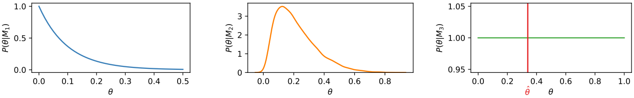
Simulation-based inference with summary statistics

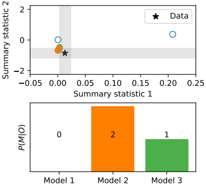




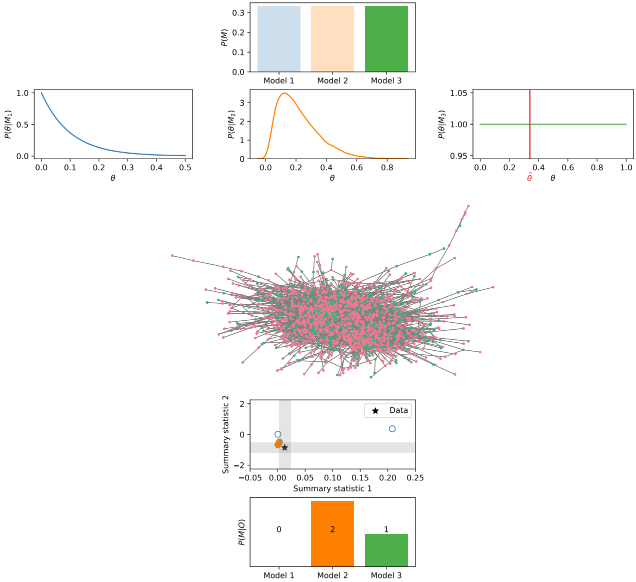
Simulation-based inference with summary statistics



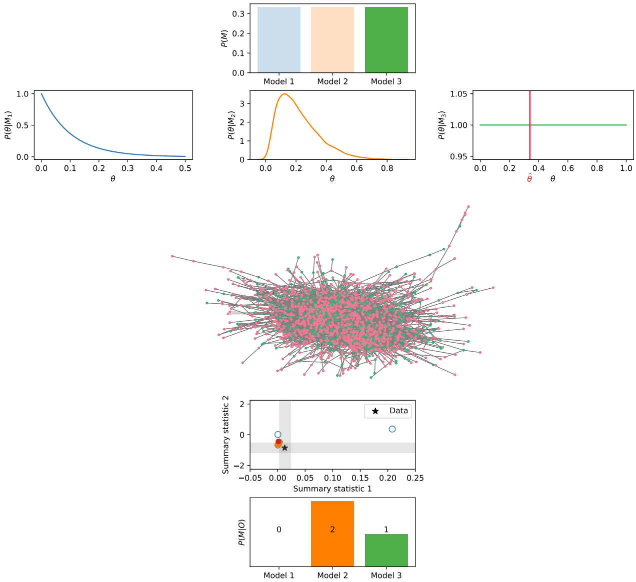




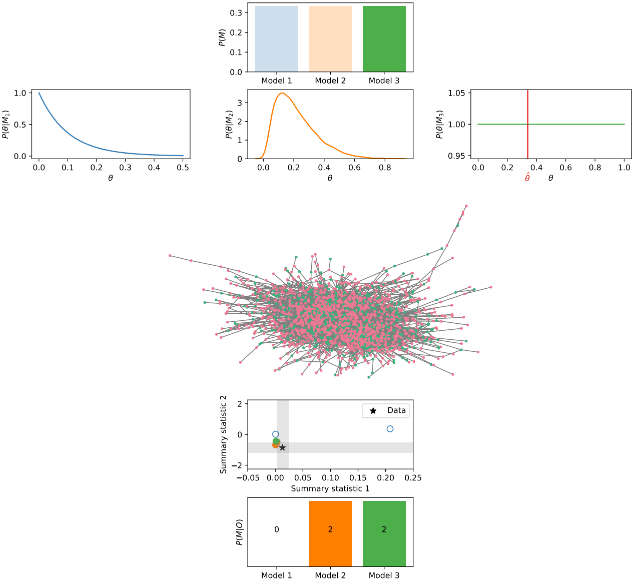
Simulation-based inference with summary statistics



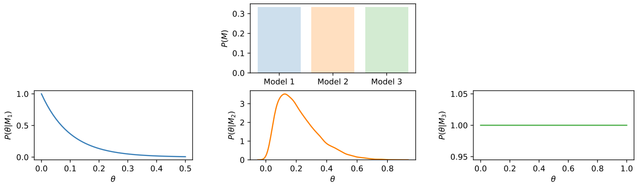
Simulation-based inference with summary statistics

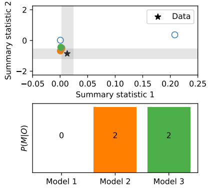


Simulation-based inference with summary statistics

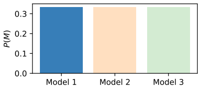


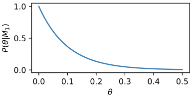
Simulation-based inference with summary statistics

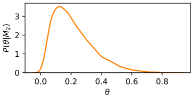


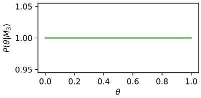


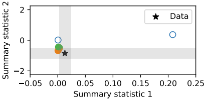
Simulation-based inference with summary statistics

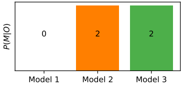




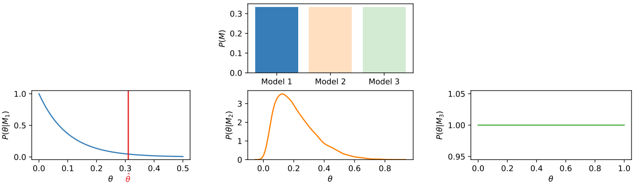


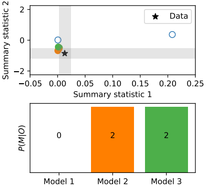




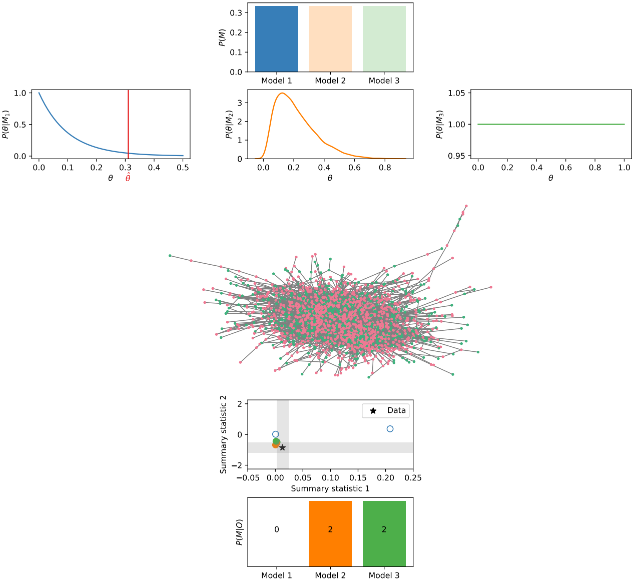


Simulation-based inference with summary statistics

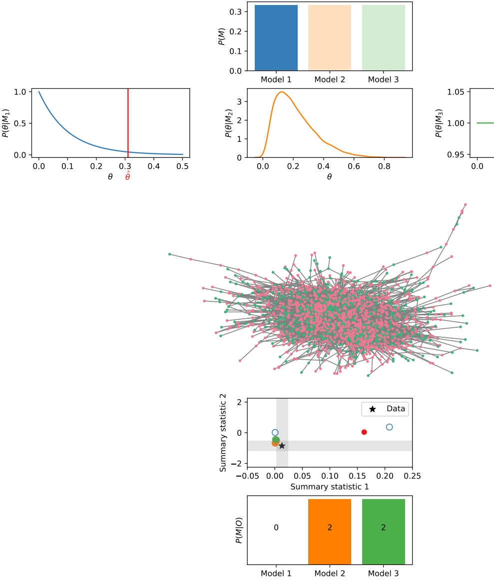




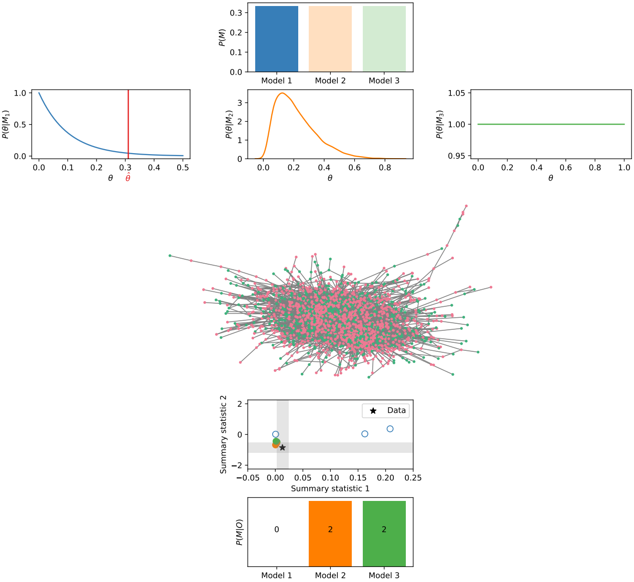
Simulation-based inference with summary statistics



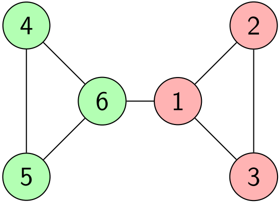
Simulation-based inference with summary statistics



Simulation-based inference with summary statistics



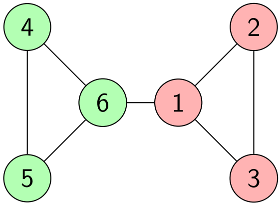
Local versus global mechanisms of coordination

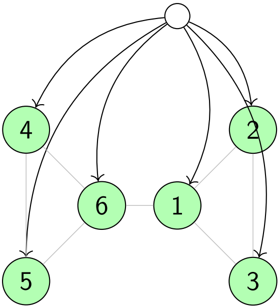


Local coordination

Strategic alignment, imitation of peers...
$\boldsymbol{J}$

Local versus global mechanisms of coordination





Local coordination

Strategic alignment,
imitation of peers...
$\boldsymbol{J}$

Global coordination

Adaptation to research purposes,
or shared culture ("disciplinary matrix")
$\boldsymbol{B}$

The Ising model as an intermediate idealized model

- Atomic magnetic spins in a material can be in two states: $\uparrow$ $(+1)$ or $\downarrow$ (-1).
- Magnetic spins prefer to be aligned to their neighbors ($\uparrow\uparrow$ or $\downarrow\downarrow$)
- Can local interactions between spins at the microscopic level lead to macroscopic alignment?

$$P(\{\sigma_i\}|J,\boldsymbol{B}) = \frac{1}{Z(J,\boldsymbol{B})} e^{-H(\{\sigma_i\},J,\boldsymbol{B})}, \text{ and } H = -\underbrace{\sum_{i,j} J w_{ij}\sigma_i\sigma_j}_{\substack{\text{local} \\ \text{pairwise interactions}}} \underbrace{-\sum_i B_{C_i}\sigma_i}_{\substack{\text{external} \\ \text{magnetic field}}} \tag{6}$$

https://mattbierbaum.github.io/ising.js/

Inverse Ising problem: $P(J, J^{\text{cit}}, \boldsymbol{B}|\{\sigma_i\})$

Local coordination in multi-layered graphs

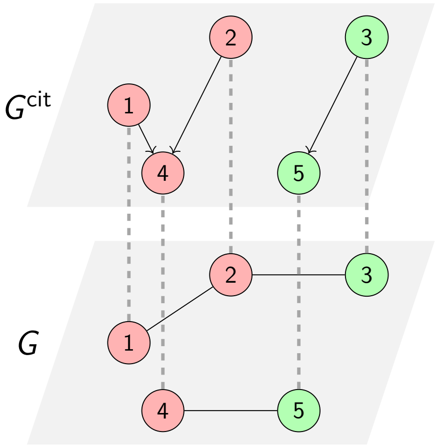


Figure: **Illustration of local coordination in multilayered social networks**. Nodes can be connected through different kinds of relationships (for instance, authors can be related via collaborations (G) or citations (G^{cit})).

Local versus global coordination

Table: Parameters of the Ising model.

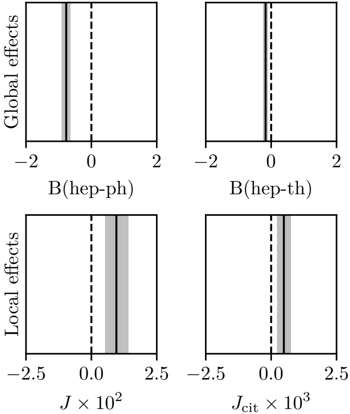
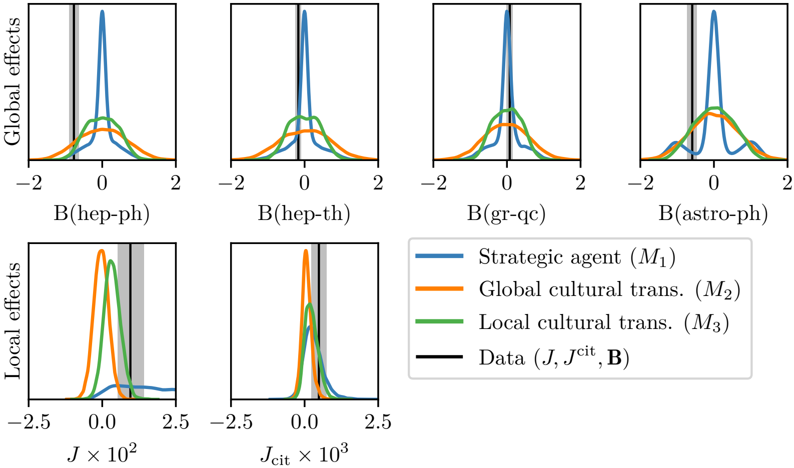


Figure: Ising model fit

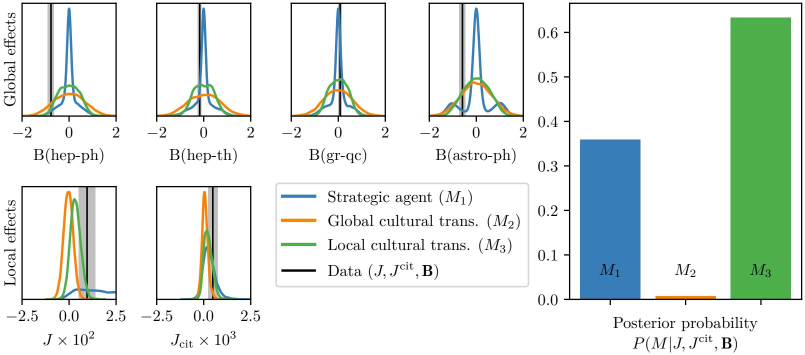
Local versus global coordination

What values of $\boldsymbol{J}$ and $\boldsymbol{B}$ do our models predict? In other words, what is the probability $P(J, J^{\text{cit}}, \boldsymbol{B}|M_i)$ for each model M_i?



Local versus global coordination

Given $P(J, J^{\text{cit}}, \boldsymbol{B}|M_i)$, and the true values of $\boldsymbol{J}$ and $\boldsymbol{B}$, what is $P(M_i|J, J^{\text{cit}}, \boldsymbol{B})$?
After a bit of computational trickery – "amortized simulation-based model comparison with neural networks" with BayesFlow –:



Challenges for model selection

- Model misspecification: model comparison among highly incorrect models is challenging/meaningless

Challenges for model selection

- Model misspecification: model comparison among highly incorrect models is challenging/meaningless
- Priors on models' parameter matter. A model is disadvantaged if it only is a good fit to the data for improbable parameter values.

Thank you!

Centola, Damon and Andrea Baronchelli (Feb. 2015). "The spontaneous emergence of conventions: An experimental study of cultural evolution". In: *Proceedings of the National Academy of Sciences* 112.7.

Delgado, Jordi (2002). "Emergence of social conventions in complex networks". In: *Artificial intelligence* 141.1-2.

Hawkins, Robert XD, Noah D Goodman, and Robert L Goldstone (2019). "The emergence of social norms and conventions". In: *Trends in cognitive sciences* 23.2.

Lewis, David (Jan. 1969). *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

O'Connor, Cailin (June 2020). "Measuring Conventionality". In: *Australasian Journal of Philosophy* 99.3.

Pujol, Josep M et al. (2005). "The role of clustering on the emergence of efficient social conventions". In: *Proceedings of the 19th international joint conference on Artificial intelligence.*

Radev, Stefan T et al. (2021). "Amortized bayesian model comparison with evidential deep learning". In: *IEEE Transactions on Neural Networks and Learning Systems* 34.8.

Inverse problems in practice

1. What **phenomenon**? (Belief-polarization? Discrimination and marginalization? etc.)
2. What **models**?
3. What **data**?
   - Accessibility (reasonable time/financial cost)
   - Quality (absence of bias)
   - Quantity (statistical significance)
4. What **computational methods**?
   - Pre-processing: text-classification? (Natural language processing)
   - Inference: simulation-based inference (with, or without neural networks)? Hamiltonian Monte-Carlo? Metropolis?

Amortized simulation-based inference

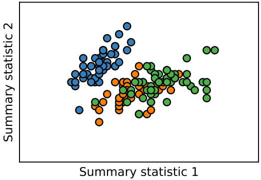
- Even with summary statistics, simulation-based inference is difficult because no simulated sample will *exactly* match the observed data.
- Solution:

Amortized simulation-based inference

- Even with summary statistics, simulation-based inference is difficult because no simulated sample will *exactly* match the observed data.
- Solution:
  - Use amortized inference with neural networks $\Rightarrow$ train a neuralnet to predict the probability of each model M_i given one or more observed outcomes. The neuralnet is trained with many simulated training samples (M_s, O_s) (Radev et al., 2021)

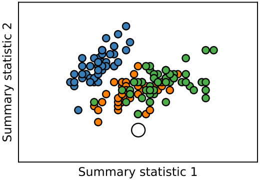
Amortized simulation-based inference

- Even with summary statistics, simulation-based inference is difficult because no simulated sample will *exactly* match the observed data.
- Solution:
  - Use amortized inference with neural networks $\Rightarrow$ train a neuralnet to predict the probability of each model M_i given one or more observed outcomes. The neuralnet is trained with many simulated training samples (M_s, O_s) (Radev et al., 2021)



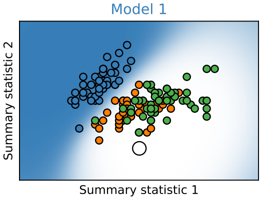
Amortized simulation-based inference

- Even with summary statistics, simulation-based inference is difficult because no simulated sample will *exactly* match the observed data.
- Solution:
  - Use amortized inference with neural networks $\Rightarrow$ train a neuralnet to predict the probability of each model M_i given one or more observed outcomes. The neuralnet is trained with many simulated training samples (M_s, O_s) (Radev et al., 2021)



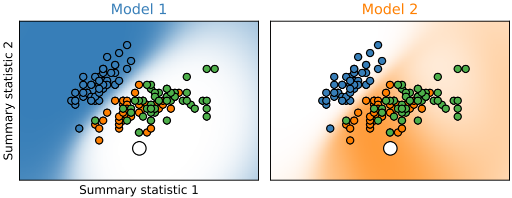
Amortized simulation-based inference

- Even with summary statistics, simulation-based inference is difficult because no simulated sample will *exactly* match the observed data.
- Solution:
  - Use amortized inference with neural networks $\Rightarrow$ train a neuralnet to predict the probability of each model M_i given one or more observed outcomes. The neuralnet is trained with many simulated training samples (M_s, O_s) (Radev et al., 2021)



Amortized simulation-based inference

- Even with summary statistics, simulation-based inference is difficult because no simulated sample will *exactly* match the observed data.
- Solution:
  - Use amortized inference with neural networks $\Rightarrow$ train a neuralnet to predict the probability of each model M_i given one or more observed outcomes. The neuralnet is trained with many simulated training samples (M_s, O_s) (Radev et al., 2021)



Amortized simulation-based inference

- Even with summary statistics, simulation-based inference is difficult because no simulated sample will *exactly* match the observed data.
- Solution:
  - Use amortized inference with neural networks ⇒ train a neuralnet to predict the probability of each model M_i given one or more observed outcomes. The neuralnet is trained with many simulated training samples (M_s, O_s) (Radev et al., 2021)