

Workshop Program Formal Models of Scientific Inquiry

Wednesday, January 29th

Time	Speaker	Title
14:00		<i>Registration and Welcome</i>
14:30	Keynote: Lorenzo Casini	Meta-Analyses and Conflict of Interest: A Bayesian Perspective
15:30		<i>Coffee Break</i>
16:00	Kevin Zollman	Authorship Norms and Epistemic Public Goods
16:35	Christoph Merdes	Gaming Testimony. An Exploration of Interest-Driven Opinion Dynamics
17:10	Matteo Micheleni	Fast, Frugal, and Effective: How Biased and Lazy Deliberation Fosters Epistemic Success
17:45		<i>End</i>

Thursday, January 30th

Time	Speaker	Title
10:00	Keynote: Ulrike Hahn	Argument in Agent-Based Models
11:00		<i>Coffee Break</i>
11:30	Martin Justin and Borut Trpin	Coherence as a Filter for Misleading Information - an Agent-Based Exploration
12:05	Edoardo Baccini (joint work with Zoé Christoff, Ludi van Leeuwen, and Rineke Verbrugge)	Learning the Truth by Selecting Good Data: Coherence Measures and Data Collection
12:40		<i>Lunch Break</i>
14:30	Keynote: Finnur Dellsén (joint work with Borut Trpin)	Testing IBE from Uncertain Evidence
15:30		<i>Coffee Break</i>
16:00	Bele Wollesen and Shira Ahissar	Mistrust of Scientific Evidence Within Marginalized Groups
16:35	Leon Assaad	Multi-Option Polarization: Modeling and Measurements
17:10	Lilian von Bressensdorf	Evidence Amalgamation: Which Aggregation Methods are Most Accurate, and When?
17:45		<i>End</i>
19:30		<i>Dinner at Ymmas</i>

Friday, January 31st

Time	Speaker	Title
10:00	Keynote: Toby Handfield	Strategic Dilemmas in Collective Inquiry: When Truth-Seeking Conflicts with Belief Revision
11:00		<i>Coffee Break</i>
11:30	Rafael Fuchs	The Zetetic Foundations of Objective Bayesianism - Efficient Inquiry and Probabilism
12:05	Klee Schöppel	Industry actors influencing science: A NormAN model
12:40		<i>Lunch Break</i>
14:30	Max Noichl	Towards Empirical Robustness in Network Epistemology
15:05		<i>Coffee Break</i>
15:40	Dominik Klein	Who Gets It Right? – On the Epistemic Performance of Democratic and Autocratic Decision-Making Procedures
16:15	Hein Duijf	Diversity and expertise in binary decision problems
16:50		<i>End</i>

Abstracts

Meta-Analyses and Conflict of Interest: A Bayesian Perspective

Lorenzo Casini, joint work with Matthias Troffaes, Jürgen Landes, and Ulrika Sahlin

Meta-analyses (MAs) are essential tools in medical research, aggregating evidence from multiple studies to inform public health and clinical decision-making. However, conflicts of interest (CoIs), such as industry sponsorship, present a significant challenge by potentially introducing bias into reported findings. Building on earlier work that used a Bayesian framework to elucidate the impact of CoIs in single-study scenarios, this paper extends the approach to MAs. With the aid of a case study on antidepressant research, we show the potential advantage of using information on CoI to revise MA estimates, in line with the Carnapian principle of total evidence, and contrary to the current protocol for MA methodology by the Cochrane Institute, which recommends reporting CoI but not adjusting for it.

Authorship norms and epistemic public goods

Kevin Zollman

Abstract: Different academic fields have different norms about how authors are listed in published papers. Some norms are insensitive to the contribution of the authors, listing them in alphabetical order or by seniority. Other norms represent the relative contributions of the authors by listing those who contributed the most first. In this paper, we develop a game theoretic model to explore the conditions under which these norms evolve and the consequences they have for collaboration. We find surprising conclusions about how the distribution of expected contribution affects the evolution of these norms. In addition, we find that all norms result in some inefficiency: they discourage some productive collaboration. Different norms discourage different types of collaboration. We explore what might be the epistemic consequence of these differences.

Gaming Testimony. An Exploration of Interest-Driven Opinion Dynamics

Christoph Merdes

Testimony has been conceived in both formal and non-formal social epistemology primarily as a decision problem: For an audience to decide how to integrate the information provided by a testimonial report provided by a partially reliable source with their credal state. This is a fruitful paradigm for many problems, but actual testimony often also contains a strategic component. Both parties are agents and may anticipate each other's behavior and divergent epistemic and non-epistemic ends. In this talk, I present a first exploratory model to include a model of strategic behavior within an opinion dynamics framework. The opinion update model is a standard Deffuant-style model, but the agents communication behavior is prior to the update governed by a game-theoretic model. IN a first exploration, I show that this model can exhibit an internal radicalization dynamic due to strategically provided deceptive reports. Further-

more, I compare convergence dynamics between a model variant employing a Chicken vs. a Prisoner’s Dilemma Game. Finally, I lay out a few important limitations of the framework as well as its potential and possible extensions and variations.

Fast, Frugal, and Effective: How Biased and Lazy Deliberation Fosters Epistemic Success

Matteo Michellini

While scholars widely acknowledge that human reasoners are biased and intellectually lazy, the effects of these traits on group deliberation remain unclear. In this work, I address this gap by developing a formal argumentation-based model to compare the epistemic performance of groups of biased and lazy reasoners with that of ideal ones. My findings show that, although ideal reasoners excel in hypothetical scenarios with unlimited time and resources, biased and lazy reasoners often achieve better outcomes under real-world constraints. I argue that bias and laziness can function as Mandevillian virtues—traits that hinder individual reasoning but enhance group performance under certain conditions. Finally, I explore the implications of these results for reasoning education and debates about the evolutionary role of reasoning.

Argument in Agent-Based Models

Ulrike Hahn

The talk will start from the multiple meanings of the term ‘argument’: as a (single) ‘reason’, a premises plus a conclusion, or a dialectical exchange. It will use these to set out a conceptual space for possible models of argument exchange in agent-based models with these three notions as axes. It will illustrate the implications and dependencies of modelling choices along these dimensions with respect to a recent Bayesian framework for modelling argument exchange across networks, NormAN (Assaad et al., 2023). It will also demonstrate how this conceptual space enables deeper comparisons between the wealth of frameworks used for modelling argument and opinion dynamics, and the implications of modeller choices among them. Crucially, it highlights not just relative strengths and weaknesses of different approaches but also important gaps in what agent-based modelling has examined.

Coherence as a Filter for Misleading Data: An Agent-Based Exploration

Martin Justin, Borut Trpin

It seems that considerations of how scientific evidence and theories fit together play a role in scientific practice. But should they? Goldberg and Khalifa (2022) have recently argued that the role of coherence in science should first and foremost be of a social and negative character. In their view, if a scientist’s beliefs do not cohere with the justified beliefs of the rest of the scientific community, then this provides a reason against the scientist’s beliefs. Proceeding from this insight that epistemic coherence in science may fruitfully be assessed in a social context, this contribution aims to provide a fuller account of coherence’s impact

on scientific communities’ epistemic performance. To achieve this, we develop a novel agent-based model in which agents try to form true beliefs about the world—represented as a Bayesian network—based on the evidence they gather. Agents can also share the evidence and, most importantly, discard it if it doesn’t cohere with their prior beliefs. We use this model to explore the value of coherence for a scientific community in various circumstances. Specifically, we are interested in situations where scientists receive misleading evidence. We ask whether coherence can help them filter such evidence and form more accurate beliefs about the world. Our results show that coherence can provide this epistemic benefit, but only to some extent, depending on the type and the amount of misleading evidence. We also explore the robustness of these results under different measurements of coherence, communication structures between agents, group sizes, and agents’ priors.

Learning the Truth by Selecting Good Data: Coherence Measures and Data Collection

Edoardo Baccini, joint work with Zoé Christoff, Ludi van Leeuwen and Rineke Verbrugge

In formal epistemology, a variety of probability-based coherence measures have been proposed that provide a quantitative formal representation of the coherence of a set of pieces of information. Research has long focused on whether coherence measures are truth-conducive. The truth-conduciveness of different coherence measures has so far been evaluated in static settings: Coherence provides assessments about the truth of incoming information but does not actively guide decisions to believe or discard pieces of information. In this paper, we propose to assess the truth-conduciveness of a coherence measure with respect to its ability to select true information and form correct beliefs in a dynamic iterated setting. At every time step, the coherence measures receive a number of noisy signals about the actual truth value of a finite set of atomic propositional random variables. By repeatedly picking the most coherent among those signals, the measures try to learn the correct truth-value assignment for the atomic variables. Our contributions are two-fold: We propose a discrete-time computational model to assess the truth-tracking abilities of different coherence measures; We compare via computational simulation a number of widely discussed coherence measures from the novel standpoint of our iterated data-collection setting. Our analysis reveals that the Glass-Olsson relative overlap measure outperforms all other tested measures when signals are not too noisy. As signals degrade, all measures become progressively worse at tracking the truth and their performance depends on the type of data collected in the first few steps.

Testing IBE from Uncertain Evidence

Finnur Dellsén and Borut Trpin

This paper investigates, through the use of computer simulations, how Inference to the Best Explanation (IBE) can be adapted to handle uncertain evidence. It focuses on evaluating IBE_{ER} , a recently proposed generalization of

IBE designed for scenarios in which the evidence is uncertain, alongside several natural alternatives to this rule. IBE_{ER} works by distinguishing between certain and uncertain evidence, making inferences only to claims that would follow from certain evidence plus some subset of uncertain evidence. The study employs computer simulations where agents apply different inferential strategies to randomly generated scenarios, receiving penalties for incorrect inferences and rewards for correct inferences (with greater rewards for more informative conclusions). The paper makes two main contributions: a systematic computational evaluation of IBE_{ER} 's performance across various conditions as compared to natural alternatives, moving beyond reliance on intuitive or historical examples; and the development of a general framework for modeling and evaluating different approaches to inference from uncertain evidence, which accommodates both probabilistic and non-probabilistic approaches.

Mistrust of Scientific Evidence Within Marginalized Groups

Bele Wollesen and Shira Ahissar

Mistrust in science among marginalized groups, such as regarding vaccine hesitancy or economic policies, is often viewed as counterproductive. However, this mistrust can arise from structural issues, notably the underrepresentation of marginalized groups in scientific research. Such exclusion can render findings less applicable to these groups, thereby justifying their skepticism. We propose a formal framework to understand how mistrust, grounded in exclusionary knowledge production, can exacerbate harm for marginalized communities. A novel variation of the two-armed bandit model explores cases where the optimal treatment for the dominant group differs from that for the marginalized group—for example, differing success rates in medical treatments across genders or differing economic outcomes between social classes. Using agent-based models with a robust set of parameters, we observe that asymmetries in information flow between the two groups, which have been prominently taken to give the marginalized group an epistemic advantage, become highly disadvantageous for that group. Additionally, we observe that, if directed towards the dominant group, higher levels of mistrust within the marginalized group improve its relative epistemic stance. This result can explain why marginalized groups tend to distrust scientific results, assuming it signifies results relevant to the dominant group rather than for them. However, we show that if mistrust is misplaced, that is, if there is general mistrust based on differences in beliefs, then the entire community does worse. In such cases, the marginalized group fares significantly worse, highlighting the dangers of misplaced mistrust and offering a potential explanation for many of the harmful societal phenomena we observe today.

Multi-Option Polarization: Modeling and Measurements

Leon Assaad

Formal models from social epistemology aim to capture and analyze polarization as the outcome of deliberation among boundedly rational agents. Classical models focus on conversations about a central proposition being true or false (Hegselmann, Krause, et al. 2006; Olsson 2013) or which of two strategies is

better (Zollman 2007; O’Connor and Weatherall 2018). Hence, agents typically face a choice between two options and may polarize accordingly. However, many real-world conversations take a different form: Which of $n > 2$ multi-option topics. This paper addresses two challenges: first, it develops a model of rational deliberation on multi-option topics. Second, it investigates whether increasing the number of discussed options (n) leads to more or less polarization. The model extends Assaad and Hahn’s 2024 simulation of a population deliberating a topic by gathering and exchanging evidence (a “NormAN” model, Assaad et al. 2023). It represents the topic as a variable with n mutually exclusive and jointly exhaustive “truth values,” allowing the number of options to vary. Do more options increase polarization? This depends on how we define multi-option polarization. Two measures are proposed: (1) the number of “positions” the population converges on, capturing the variety of favored options, and (2) belief dispersion, quantifying disagreement about each option. Preliminary results show that increasing n both exacerbates and reduces polarization: the number of defended positions rises, making it harder to settle on a single preferred candidate, while average belief dispersion decreases, reflecting less extreme disagreement about individual options. While our simulations suggest a clear trend, they raise essential questions about the measurement of multi-option polarization.

Evidence Amalgamation: Which Aggregation Methods are Most Accurate, and When?

Lilian von Bressensdorf

Stegenga (2013) highlights a fundamental challenge in evidence-hypothesis relations: it is often impractical or impossible to determine the precise likelihood $P(E_i|H)$ of a piece of evidence E_i given a hypothesis H . Since Bayesian reasoning depends on these likelihoods, absolute confirmation measures are often unavailable. Instead, evidence may only provide ordinal confirmation orderings of hypotheses. This limitation poses a significant problem: without exact likelihoods, traditional Bayesian updating cannot aggregate evidence. Aggregation rules from social choice theory are needed to combine multiple ordinal rankings into a single coherent ranking. Yet, the truth-conduciveness of aggregation rules, as explored by Hartmann and Sprenger (2012) and Volzhanin et al. (2017), remains unresolved. This project addresses two core questions: (a) Which aggregation rules best approximate optimal Bayesian reasoning with ordinal evidence? (b) Under what conditions can aggregation rules be considered effective? We propose a simulation model where an agent ranks N competing hypotheses using M evidence pieces, each with n possible states corresponding to hypotheses. Evidence reliability ($r \in [0, 1]$) modulates how accurately the evidence supports the true hypothesis. While cardinal rankings theoretically exist, only ordinal rankings are available to the agent. Consequently, the agent relies on aggregation rules from social choice theory rather than Bayesian reasoning. Preliminary findings show differences in aggregation rule performance. While higher reliability and more evidence improve identifying the single true hypothesis, they do not enhance ranking accuracy across all N hypotheses.

Strategic Dilemmas in Collective Inquiry: When Truth-Seeking Conflicts with Belief Revision

Toby Handfield

Many forms of collective inquiry involve a tension between individual and group epistemic goals. One tension arises when individual agents can identify false possibilities within a shared hypothesis space, but face personal costs in revising their own beliefs. Examples include scientific research programs eliminating candidate theories, forensic teams ruling out suspect hypotheses, and informal discussion and debate. While the collective benefits from rapidly identifying truth through elimination of falsehoods, individuals may prefer that others do more of the work of exploring and rejecting possibilities. We develop a model to study the strategic interactions arising from inquirers with two motivations: seeking truth and minimizing changes to their own beliefs. Our framework predicts social dilemmas where individually optimal strategies - such as waiting for others to do the investigative work - lead to suboptimal collective discovery of truth. We further show that, paradoxically, increasing agents' motivation to seek truth can lead to less favorable equilibrium outcomes for truth discovery. These results offer insights into the dynamics of collective inquiry, highlighting how even individuals with relatively pure epistemic motives can sometimes fail to achieve optimal epistemic outcomes.

The Zetetic Foundations of Objective Bayesianism: Efficient Inquiry and Probabilism

Rafael Fuchs

In scientific inquiry, efficient use of resources is of paramount importance: experiments can be costly in terms of time and materials, while budgets are limited. Therefore, it is in the interest of scientists (and society) that inquiry is performed 'efficiently', i.e. without an unnecessary waste of resources. In this paper, I present a formal model of efficient inquiry, and explore its implications. The starting point is a scientist (or group of scientists) trying to figure out the true value of some variable X of interest. To this end, they can choose from different sequences of experiments (called "inquisitive strategies") that successively narrow down, and eventually single out the true value of X . When choosing an inquisitive strategy, scientists need to take prior evidence and their rational beliefs into account, in order to avoid starting with a line of inquiry that they can already expect to yield no informative results. Given that scientists want to avoid such inefficiencies, it turns out that their beliefs need to satisfy the axioms of probability (i.e. "probabilism"). Specifically, credences that violate the axioms of probability will be susceptible to choosing strategies that contain inefficiencies, which probabilistic credences will avoid. Importantly, the requirement of (imprecise) probabilism follows not only when numerical credences are presupposed, but also when we start from comparative (i.e. ordinal) belief relations. Finally, the model provides insights into the design of efficient experiments, and how to measure their information content.

Industry actors influencing science: A NormAN model

Klee Schöppel

As documented in [1], political and societal consensus on the perils of human-made climate change has decreased since the 1990s, as a direct result of intentional, ideologically motivated propping up of the notion of scientific uncertainty on the issue. The authors identify these strategies as adapted from corporate PR efforts to undermine certainty in the harms of smoking tobacco. Since then, multiple publications [2, 3, 4, 5] have provided additional insights into the social epistemological processes involved when the 'Tobacco Strategy' so disrupts evidence-based policymaking, by offering representations of the sub-processes involved (Biased production, selective sharing and industry selection) in the form of agent-based Bala-Goyal models. Inspired by concerns about the suitability of Bala-Goyal models to the representations of scientific inquiry above and beyond specific research questions in the medical context (such as the peptic ulcer case study in [6]), I propose an alternative model of the Tobacco Strategy built instead on the NormAN-framework [7]. Therein, a scientific community investigates a hypothesis under consideration based on a limited number of distinct, causally relevant pieces of evidence, which may be shared as parcels across their network. I implement the Tobacco Strategy's impacts on the methods and directions of agents' inquiry and combine them with selection rules for the communication of evidence. My model strengthens the existing case for industry funding of research depending, in its effectiveness, directly on its ability to exploit competitive structures in funding allocation to academic research, even in the absence of outright fraud by individually compromised researchers.

1. Oreskes & Conway (2010) Merchants of Doubt.
2. Holman & Bruner (2017) Experimentation by Industrial Selection.
3. O'Connor & Weatherall (2018) The Misinformation Age.
4. Holman, Bennett & Bruner (2015) The Problem of Intransigently Biased Agents.
5. Weatherall, O'Connor & Bruner (2018) How to Beat Science and Influence People.
6. Zollman (2009) The Epistemic Benefit of Transient Diversity.
7. Assaad, Fuchs, Jalalimanesh, Phillips, Schöppel & Hahn (2023) A Bayesian Agent-Based Framework for Argument Exchange Across Networks.

Towards Empirical Robustness in Network Epistemology

Maximilian Noichl, joint work with Ignacio Quintana and Hein Duijf

One of the central results in simulation studies of science argues that less communication often leads to higher reliability. More generally, simulation studies have been used to explore which communication networks enhance collective reliability and speed of convergence. However, this literature has largely concentrated

on relatively simple network structures (e.g., cycles, wheels, full graphs), which bear little resemblance to real social networks. This has led to worries about the robustness of these effects: Does less communication lead to higher reliability in realistic social networks? In this talk, we provide the first results concerning the empirical robustness of these results with respect to real social networks. We develop a novel method to perform this empirical robustness analysis. First, we use citation graphs to depict empirical networks commonly discussed in the literature as examples of lagging discovery—i.e., one concerning the bacterial causes of peptic ulcers and another concerning the prolonged history of the perceptor. Second, we develop a new method to generate a sample of "similar" networks. Third, we work out the collective reliability and speed of convergence of these communication networks by running simulations. Finally, we analyse the data about these networks to determine which outcomes can be expected in real networks and which network properties (e.g., connectivity, degree inequality) strongly affect collective reliability and speed of convergence.

Who Gets It Right? – On the Epistemic Performance of Democratic and Autocratic Decision-Making Procedures

Dominik Klein

How can governance systems achieve good collective decisions? We examine the epistemic foundations of good governance within democratic and autocratic systems, contributing to debates on the epistemic justification of democratic procedures (Estlund 2000, Landemore 2013, Goodin 2018). Using a simulation model, we analyze a thought experiment where agents estimate the optimal level of public good provision. Our findings reveal three key insights: i) democratic decision-making outperforms autocratic methods in judgment accuracy, ii) democratic systems excel when citizens display a mild bias toward their own needs rather than complete impartiality, and iii) limiting deliberation time can, under specific conditions, enhance the epistemic precision of decisions. These insights underscore the epistemic strengths of democracy in navigating complex governance challenges.