MBR026_ROME
Tenth International MBR Conference

MODEL-BASED REASONING

Epistemology, Artificial Intelligence, Cognitive Science

17-19 June 2026 - Rome, Italy

Book of Abstracts

(in alphabetical order for first authors)

SAPIENZA
UNIVERSITÀ DI ROMA

UNIVERSITÀ
DI PAVIA

Keynotes

Bertolaso Marta – **Shared Agency in Model-Based Reasoning**

One Shared agency (SA) has emerged as a pivotal conceptual construct in the modelling of Human-Robot Interaction (HRI). The framing of the systems and the models setting, however, are arising different epistemological concerns that span from explanatory issues to models and theories compatibilities.

In the first part of the paper, I present a systematic review of three dominant perspectives that structure current scientific practices in the HRI field.

The Cognitive Perspective enquires agency primarily in terms of goal alignment: SA is understood as the congruence between predicted and observed sensory outcomes, where reciprocity in HRI is modelled as a feedback loop oriented toward prediction-error minimisation. The Embodiment Perspective frames agency as bodily extension, the dynamic integration of robotic artifacts into the body schema, resonating with epistemological questions already discussed in the philosophical literature on systemic and context-dependent frameworks of cognition. The Engineering Perspective treats agency as a design variable to be optimised within control architectures: consistent with informational theories, error minimisation constitutes a typical success parameter, with zero error and full assistance often functioning as the design ideal, and SA operationalised as a parameter of system performance. I will argue that the most significant epistemological tension among these models emerges, however, from examining the specific role that SA plays in the different explanatory accounts. For the Cognitive and Embodiment perspectives, SA functions as an a priori explanatory tool, a conceptual framework deployed to account for feedback loops and bodily coupling in HRI. For the Engineering perspective, SA instead becomes a constitutive part of the explanandum: the specific dynamic under investigation, understood in terms of the reciprocity of interactions between humans and robots as it is actually enacted.

I will conclude that a relational epistemology, one that takes the human-robot-environment assemblage rather than any single component as its primary unit of analysis, offers a productive path for disentangling these tensions and making the most of the scientific practices they generate.

Bangu Sorin Ioan – **Rule-following and Rule-breaking: How is Scientific Creativity Possible (in the Age of AI)**

Few notions are more important and yet less understood than scientific creativity. In this talk, I aim to take some steps toward articulating a philosophical model helping us make better sense of the creative aspect of science (and mathematics). The model builds on Margaret Boden's distinction between combinatorial and transformational creativity (Boden 2004; 2014), and integrates Feyerabendian insights about his so called 'epistemological anarchism', as well as the 'paradox' of rule-following (that Kripke claimed to find in Wittgenstein's Philosophical Investigations). However, I propose somewhat different accounts of both the combinatorial and the transformational kinds of creativity, where these modifications are suggested by reflections on several episodes in the history of science (and mathematics). Importantly, the model doesn't presuppose (nor entails) that creativity is an exclusively human ability, so it allows us to consider the possibility, and hope, that AI systems (the current LLMs or of other type) could, and will, find creative solutions to outstanding scientific problems. I shall argue that although the model doesn't apriori rule out this possibility, and hope, it also gives us a sense of how we should recalibrate such expectations.

Gelfert Axel – **When Models Trespass**

Models are ubiquitous in science; they function as both representational devices and epistemic tools. We routinely turn to them for scientific understanding and answers to scientific

questions. While it would be fanciful to attribute agency to them, we often treat them as informants on complex matters, and like informants they are reliable only within a certain domain. Problems arise when models trespass beyond their legitimate domain of application. In this keynote, I will outline conditions for when models trespass and highlights some of the epistemic pitfalls and non-epistemic risks associated from such an overextension of models.

Ladyman James – **The Limits of Pluralism and Realism**

Models are the subject of a vast literature in philosophy of science, and there are different accounts of their nature, and how they represent real systems, and different accounts of scientific representation in general (Hesse 1963, Cartwright 1983, vanFraassen 2008, Frigg and Nguyen 2020). Some philosophers deny or downplay the representational role of models (for example, Knuuttila 2003 who regards them as epistemic artefacts). Van Fraassen argues that models only represent given an agent with a purpose. Nonetheless it is widely agreed that models are sometimes representations of real phenomena and systems, even if that is not their primary role, and even if they do not represent independently of being used to do so if at all. Furthermore, there is much more that can be said about models and scientific representation about which philosophers of science agree, or should agree according to this paper. One crucial issue that still divides realists and antirealists is the whether models represent causal structure or mechanisms. This paper considers the implications for pluralism and realism of the fact that successful modelling involves empirical adequacy that is always to some degree of precision and always involves phenomena in some domain or at some scale.

Magnani Lorenzo – **Beyond Stochastic Parrots. Large Language Models as Superior Masters of the Symbolic: The Jeopardy of Human Creativity in the Age of Cynical AI Ethics**

This intervention builds on the paper Enhancing or Jeopardizing Human Creativity? Will Humans Be Able to Defend Themselves Against AI Superpowers in an Age of Ethics Washing and Law Washing? (Magnani, 2026) and examines the epistemic threats posed by Large Language Models (LLM models). It asks whether human creativity can still be protected when these AI systems, now acting as "Superior Masters of the Symbolic", are deployed under the dominant regime of "ethics washing" and "law washing"

The paper argues that academic AI ethics almost always ideology in its purest, most cynical form today. It is not hypocrisy; it is cynical reason – we all know very well that the principles have no teeth, yet we continue to act as if they matter. This cynical reason, in the age of ethics and law washing, produces only performative moral theater while corporations and states race ahead without any real acts of a "violent arm of the law". The emptiness of the discourse is not accidental — it is the function of the system perpetrated by corporations and submissive or ineffective parliaments. It keeps the system lubricated.

Drawing on my eco-cognitive framework, the talk shows how LLM models, as Superior Masters of the Symbolic, risk turning human reasoners into passive consumers of pre-fabricated inferences, thereby jeopardizing human specific abductive creativity. It wraps off by describing defensive tactics based on a revitalized epistemology that can take on AI superpowers.

Palacios Patricia – **Modeling Instability: From Eological Collapse to Political Transitions**

Minor perturbations can occasionally set off cascading effects that lead to abrupt shifts in ecological and political systems; such shifts are often described as ecological and political tipping points. Examples are the collapse of fishery in coral reefs caused by the addition of a small number of predator fish or political revolutions triggered by apparently minor factors, such as the increase of a few cents in the metro ticket. Although these shifts are extremely hard to predict and cannot be fully captured by mechanistic models, some characteristics consistently emerge as indicators of the kinds of instabilities leading to abrupt transitions. While some indicators, such as positive feedback and volatility are more generic, others, such

as preference falsification and homophily, are specific of sociopolitical systems with no analogue on ecological systems. In this talk, I will stress the analogies and disanalogies between ecological and political transitions and the consequences for model building. I argue that the most promising way to apply mathematical theories of ecological tipping points —such as bifurcation theory or theories of criticality—to the political case is to construct new models that integrate universal features of critical transitions with features that are distinctive to the political domain. This, in turn, will guide us toward a new framework of model transfer, which I call "model adaptation."

Portides Demetris – **The Function of Abstractions and Idealizations in Scientific Models and Scientific Understanding**

Several pragmatically oriented approaches to the notion of scientific understanding rest on the observation that scientific models are constructed using abstractions and idealizations and since the latter deviate from truth, the models they give rise to lead to false claims about worldly systems. Since, as it is claimed, false models cannot contribute to truth but do facilitate scientific understanding, scientific understanding is not necessarily connected to truth. A robust version of this argument leads to the claim: Since model representations of wordly systems are ubiquitous in scientific practice and since they cannot lead to truth, scientific understanding is a primary scientific aim that displaces or outplaces the scientific aim of truth.
I raise two concerns with the robust version of the argument on grounds that I shall motivate in my talk. The first relates to how the representational relata are conceived and the second relates to the inclusion of abstractions and idealizations within the general category of "falsehoods" without further qualification. I briefly elaborate on the first concern here. The robust pragmatic argument is based on an understanding of the relata as follows: a theoretical scientific construction, i.e. a model, is meant to represent a particular concrete target system in the world. If such were the case, then the argument for displacing truth could be sound. But, I think the facts of the matter are far more complex. Scientific models involve different levels of generality. For instance the linear harmonic oscillator, LHO, is meant to represent only the effects of a linear restoring force due to gravity. To claim that LHO is false, because it does not represent accurately and precisely the behavior of an actual oscillating body, is to overlook the purpose of the model. Scientific models most often represent theoretical types (or generalized phenomena), and their truth-valuation is not something that could be determined in a straightforward way by correspondence to the world. Truth in science is relative to the level of generality of models. Therefore, the claim that the epistemic aim of scientific understanding outplaces the epistemic aim of truth is farfetched.

Schiaffonati Viola – **Towards the New Sciences of the Artificial: The Role of Patrick Suppes' Philosophy of Science**

One of the key roles of philosophy of science is dealing with the philosophical foundations of scientific disciplines, that is to elicit and analyse the basic notions and concepts at the roots of a discipline. In this talk I will focus on the foundations of the artificial sciences (notably Artificial Intelligence but not only) inspired by some ideas of Patrick Suppes' philosophical approach. First, I will discuss why we need to lay down the foundations of new sciences of the artificial, partly in continuity and partly in discontinuity with Herbert Simon's seminal work. Second, I will outline some of the overarching themes of Suppes' philosophy of science and use them to set the foundations of the new sciences of the artificial in a programmatic way. Third, I will illustrate the profitability of taking inspiration in this endeavour from Suppes' contribution. While Suppes' work has shaped much of current philosophy of science, his contributions are not often discussed. In this respect, my talk aims at illustrating and concretely using them in setting out the research agenda of the new sciences of the artificial. In particular, I will focus on Suppes' bottom-up approach and his pluralist and pragmatist stances; on the centrality of models and their hierarchies; on the role of uncertainty and the crisis of the universality of methodological standards.

Papers

Addis Mark; Claudia Eckert – **Developing Explanatory Frameworks in Complex Model Based Reasoning Systems Using Category Theory**

Complex socio-technical systems are heterogeneous so they cannot be modelled by a single approach (Simon 1962). From a socio-technical systems modeling perspective, what is required is a commitment to the possibility that a given model and its relationships with other models can be fictional (Frigg 2010). A key modeling objective is the production of models that are effective decision-making aids (Eckert and Hillerbrand 2022). The property of potentially being part of multiple subsystems makes complex systems inherently difficult to model, with model analysis being affected by decisions about the allocation of parts to models. Epistemologically adequate modeling accounts need to offer a way of addressing such divergences and potential incommensurability. Producing and combining complex heterogeneous models is theoretically and practically challenging. One approach to this is applied category theory whose basic concept is that many formally describable structures can be described as maps (Mac Lane 1998). Importantly category theory is able to both state mathematical propositions and be a mode of mathematical reasoning in flexible and complementary ways.

Complex model-based reasoning systems are only as valuable as the theoretically informative, practically reliable, ethically and legally acceptable outputs they produce. The use of artificial intelligence in such systems is increasingly common and has become crucial to the epistemology of complex systems, including what constitutes a reliable model and the nature of good explanations in these systems. Characterizing and clearly defining what data traceability means is also a central part of data traceability methodology. Category theory is particularly effective at handling the different ways data may be defined as fixed and as interpretable and in changes to these ways because it emphasizes relations and structures. Practically data traceability methodology involves finding appropriate formal representations of the agreed data semantics and characteristics using tools from applied category theory which handle the transformation of data.

References

Addis M. and Eckert C. (2024). Explanatory frameworks in complex change and resilience system modelling. Logic Journal of the IGPL. https://doi.org/10.1093/jigpal/jzae087.

Eckert, C., and Hillerbrand, R. (2022). Models in engineering design as decision-making aids. Engineering Studies, 14 (2), 134-157.

Frigg, R. (2010). Models and fiction. Synthese 172 (2), 251-268.

Mac Lane, S. (1998). Categories for the Working Mathematician. 2nd ed. Springer.

Simon, H.A., (1962). The architecture of complexity. Proceedings of the American Philosophical Society, 106 (6), 467-482.

Aguilar Omar; Anthony Aguirre – **How Are Scientific Concepts Birthed? Typing Rules of Concept Formation in Theoretical Physics Reasoning**

This work aims to formalize some of the ways scientific concepts are formed in the process of theoretical physics discovery. Since this may at first seem like a task beyond the scope of the exact sciences, we begin by presenting arguments for why scientific concept formation can be formalized. In particular, we put forward the idea that intuitive types are a central guide of scientific concept formation. Then, we introduce type theory as a natural formalization of these intuitive types. We formalize what we call "ways of discovering new concepts" including property preservation and concept change, as cognitive typing rules. Next, we apply these cognitive typing rules to a case study of conceptual discovery in the history of physics: Einstein's conceptual path to the relativity of time. Then, we recast what a physicist might informally call "ways of discovering new scientific concepts" as compositional typing rules built from cognitive typing rules, thereby formalizing them as scientific discovery mechanisms. Notably, we formalize Einstein's discovery mechanism that led to the relativity of time as the assumption-conclusion rule, which switches the assumption and conclusion of a line of

reasoning. Lastly, we computationally model the type-theoretic reconstruction as a program synthesis task and show that Neuro-Bayesian search can mitigate the curse of compositionality.

Amato Alessandro – **From Rationality to Ordered Emotionality (Affectus Ordinatus): A Framework for Human and Artificial Linguistic Generation**

This contribution reconsiders the historical evolution of the concept of rationality, from its embodied origins to its disaffectivized formalization in the twentieth century. It proposes the concept of Ordered Emotionality as a functional reformulation of the rationalist paradigm, redefining its affective presuppositions. Philophonic Logical Centering is introduced as a pre-verbal process that stabilizes affective and identity-related configurations, rendering emotional configurations logically usable in the selection of linguistic action. This process enables the Philophonic Emotional Distillate, which transforms emotion from automatic reactivity into a selective expressive resource. A pilot study (N = 8) indicates a relationship between pre-verbal stabilization and prosodic coherence. Finally, the work suggests implications for Artificial Intelligence through the integration of mechanisms of preventive affective regulation.

These findings contribute to current debates on Human-Centered Artificial Intelligence, suggesting a shift from purely predictive architectures toward affectively regulated, action-oriented systems.

Amigoni Francesco – **On the Models for Automated Planning**

Planning is a core capability of intelligent agents and involves, in its classical version, determining a course of actions, a plan, that achieves a given objective. Starting from the early days of Artificial Intelligence, planning has been studied across a wide range of settings using different methods. These techniques rely on explicit models of the environment and of the actions to forecast what happens when sequences of actions are performed, and they have been successfully applied in domains such as robotics, logistics, manufacturing, and videogames.

Recently, the rise of generative AI has promoted interest in using Large Language Models (LLMs) for automated planning. LLMs, neural networks trained on massive text corpora to predict and generate language, can be prompted with a natural-language description of an objective and asked to produce a plan as a sequence of actions. While appealing, this end-to-end approach has been shown to underperform compared to LLMs' success in other tasks. This talk argues that a fundamental reason for this limitation is that LLMs rely on weak and inaccurate implicit models of the environment and the actions. In classical planning, these models are essential: they allow an agent to predict the outcomes of actions and reason about future states. With accurate models, traditional planning methods reliably guarantee that the objective will be achieved when executing a plan. In contrast, LLMs derive their models implicitly from general-purpose text data, which makes them unreliable for forecasting action outcomes, especially outside the training distribution. Empirical evidence shows that LLM planning performance degrades as plan length increases, further indicating model inadequacy. The talk concludes by discussing whether these limitations are inherent to LLMs or whether future advances could enable them to produce provably correct and reliable plans.

Avitabile Piero; Matteo De Benedetto; Edoardo Peruzzi – **Scientific Experts: Which Questions Should We Ask?**

Scientific experts play a central role in public decision-making. Yet reliance on their judgments raises an epistemic problem: how can the reliability of expert judgments be assessed by lay decision-makers who lack the relevant technical competence? Philosophical accounts of expertise have traditionally addressed this problem by adopting an agent-centered perspective, focusing on indirect indicators of reliability such as credentials, track records, and epistemic or moral virtues. While these approaches offer valuable guidance for rational deference to experts, they tend to treat reliability as a relatively stable property of experts as epistemic agents. This paper argues that such an agent-centered conception of expert reliability is

systematically incomplete. We propose that the reliability of expert judgment is not solely a function of who the expert is, but also of what question is being asked and whether that question can be reliably answered given available epistemic resources and contextual constraints. Expert reliability, we contend, is fundamentally question-dependent. The same expert judgment may be highly reliable with respect to one problem while being poorly suited, or even misleading, with respect to another. To capture this insight, we introduce a distinction between two complementary dimensions of expert reliability: agent-oriented reliability and problem-oriented reliability. Agent-oriented reliability concerns the expert's competence, methodological rigor, and epistemic virtues, whereas problem-oriented reliability evaluates the fit between the question posed, the structure of the problem, and the methods, models, and evidence employed. This second dimension highlights the role of institutional contexts, decision requirements, and question formulation in shaping the reliability of expert judgment. We develop this framework through a critical review of agent-centered approaches in social epistemology, a conceptual analysis of problem-oriented reliability informed by methodological debates on validity, and a case study of merger simulation models in competition policy. The paper concludes by arguing that expert reliability is a co-constructed property emerging from scientific practice, institutional design, and public uptake, with important implications for policy-making and the epistemology of expertise.

Barés Gómez Cristina; Matthieu Fontaine – **Abduction, Conjectures, and Meaning**

Abduction is a form of reasoning that introduces hypotheses to address epistemic gaps. Building on Peirce's original conception and subsequent developments, abduction is here understood as a cognitively productive process. Its pragmatic dimension plays a central role in the interpretation of linguistic meaning, particularly in the generation of Gricean implicatures. Abductive conjectures, formulated and activated within language games, allow interlocutors to reconcile utterances with conversational norms in the absence of full information. This perspective situates abduction at the interface of epistemology, pragmatics, and the philosophy of language, highlighting its role as a dynamic mechanism in meaning construction. In this paper, we address the issue in the two following way: what is the meaning of abductive conjectures, and how meaning can be conjectured abductively.

Bianchini Francesco – **Analogy as Model-Based Reasoning in Generative AI Systems**

The paper analyzes analogy as a distinctive form of model-based reasoning, grounded not in simple similarity but in a relation between relations. In every analogy, one domain functions as a model (the source) for another (the target), enabling epistemic expansion either through discovery or argumentation. This conception allows a clear distinction between source and target domains, each characterized by its own internal relational structure, with the source relation serving as a model for the target relation. From this perspective, analogy has been extensively studied in contemporary cognitive science, particularly within structure-mapping theories that model analogy as the alignment of relational structures and support explicit computational implementations. Alternative approaches emphasize a more contextual, bottom-up view. Analogy has also played a central role in accounts of scientific modeling and discovery, where analogical inference relies on similarities in observable properties and the alignment of causal relations between source and target. Within artificial intelligence, analogy has been modeled through symbolic, multiconstraint, and connectionist approaches. While symbolic models render analogical structures explicit, neural-network-based models encode them implicitly in distributed patterns. In both cases, the model involved in analogy can be identified and used to explain the system's behavior. This explanatory link becomes weaker in generative AI, especially in large language models, where a growing separation emerges between the identifiable analogical structure in the output and the opaque internal processes that produce it. LLMs nevertheless exhibit remarkable analogical abilities, including the generation and evaluation of original scientific analogies. This raises the question of whether their performance reflects genuine source-target mapping or merely sensitivity to learned relational regularities. A leading hypothesis attributes their analogical capacities to high-dimensional vector embeddings, where analogy arises as a geometric transformation rather

than explicit relational manipulation. The paper ultimately aims to assess whether and how such mechanisms can be aligned with human analogical reasoning.

Bouchard Yves – **Inferential Knowledge and Object-Oriented Epistemology**

When considering a fragment of what an epistemic agent knows, one can collect this information in a set in which all epistemic items can be represented as assertions. Such assertions, in turn, can be conceived as facts or rules for asserting facts. A set that contains both facts and rules is a knowledge base. A knowledge base can be queried to verify whether an assertion about a fact is true, and such a process can be assimilated to an inference. Now, if the knowledge base is populated with assertions of different types, i.e., epistemic items of different types such as perceptual knowledge and a priori knowledge, then under which conditions exactly can one perform an inference involving different knowledge types? Can an inference operate on two different epistemic types? And if so, what is the epistemic type of the conclusion? More generally, if a theory of knowledge allows for different epistemic types, then under which conditions can knowledge be exploited by inferential means? One direct consequence of not representing explicitly epistemic types in the case of an inference is the risk to be exposed to what can be called the type-mismatch problem, i.e., the problem of drawing a conclusion whose type is incompatible with the types of the premises.

In the first part of the talk, I will analyze the type-mismatch problem for inferential knowledge and I will highlight some of its consequences on a theory of knowledge. In the second part, I will present three forms of abstraction on types (logical, procedural, functional) in order to make distinctions between different representations of types, and I will propose an object-oriented analysis of epistemic types (inspired by the object-oriented programming in computer science) capable of making explicit type matching in the context of inferential knowledge.

Bringsjord Selmer; Naveen Sundar Govindarajulu; Alexander Bringsjord – **Heterogeneous Abductive Logics**

We introduce herein a new space of (formal) abductive logics: *heterogeneous* abductive logics: *HAL*. To do so, we (1) briefly note that hitherto in the study of abductive logic, to the best of our knowledge heterogeneous logics are absent; (2) summarize the essence of heterogeneous logics, which first appeared on the formal-logic and logic-based-AI scene circa 1990; (3) characterize a particular but representative abductive logic *AHCEC* within *HAL*, by encapsulating five desiderata that we hold must be satisfied by formal heterogeneous abductive logics; (4) explain this logic with help from some examples that -- true to the nature of what a heterogeneous logic is -- require some reasoning over visual/imagistic content. (5) Next, we demonstrate some heterogeneous deductive reasoning in an AI-infused interactive proof-and argument-construction environment (HyperSlate^R). (6) We end with some brief remarks regarding the future building out of the new logics in question, which includes developing both an automated heterogeneous abductive reasoner and associated verifier.

Brinz Johannes – **What Are Neurocomputational Models?**

The idea that the brain computes is fundamental to computational neuroscience. However, the precise nature of "neurocomputational models" and the systems they describe is still debated. I argue that these models constitute a subclass of mathematical models in the natural sciences, distinguished by their medium-independence. Like other mathematical models, they represent concrete target systems with mathematical descriptions. However, unlike other mathematical models, which describe specific physical quantities measured in corresponding units, neurocomputational models abstract away from material details, interpreting their variables as unitless computational entities. More specifically, my claim is that neurocomputational models describe the way the brain processes Shannon information. In this way, neurocompuational models inherit two properties of Shannons information that are central to their understanding: First, Shannon information is an observer-independent measure of a concrete target system. Second, Shannon information is medium-independent, as system with

different phyiscal makup can process Shannon information in exactly the same way. This view reframes the relationship between computational models and brain processes from one of implementation to one of scientific representation. This proposal yields several advantages. First, it situates computational neuroscience firmly within the natural sciences. Second, it helps to distinguish brain-like (neuromorphic) computers from digital brain simulations. This has consequences for debates on artificial consciousness. Third, it differentiates between digital and analog models, by conceptualizing the former as using discrete (random) variables, and the latter using continuous (random) variables. Fourth, the proposed accound could provide the basis for a unified theory of computation, conceptualising abstract computations as mathematical models of physical computations. The idea being that not only neurocomputational models describe physical computations, but the computational models discussed in computer science (Turing machines, finite state automata, etc.) do as well.

Caterina Gianluca; Rocco Gangle – **A Minimal Reconstruction of Eg_{β} Logical Rules from Eg_{α}**

The propositional fragment α of Peirce's existential graphs (EGα) admits a proof theory of polarity-sensitive insertion (in odd regions) and erasure (in even regions), together with iteration/deiteration and the double cut. Standard presentations of the first-order fragment β, however, add a separate family of "line rules" for the line of identity (extension/retraction, transport across cuts, and join/split), treating identity manipulation as a distinct stock of inference principles. We present of reorganization of β by moving first-order content from rules into the syntax of what may be inserted, erased, or iterated. We develop the key of a "ported inscription", which is a subgraph with an interface of ports recording how identity segments meet the boundary of its region. This, in turn, supports "anchored" iteration: when a ported inscription is copied into a nested region, each port of the copy must attach to the same identity component as the corresponding port of the original. To handle changes in identity, we propose the use of "local bridges", which are defined a identity segments with two ports/hooks whose attachments are restricted to identity stubs in the same region, preventing propagation across cut boundaries. Under locality, join/split become insertion of a bridge in an odd region and erasure in an even region. Finally, we prove soundness for β semantics and show each standard β rule is derivable in the ported system, hence equivalence on closed graphs.

Cecchet Matteo – **Model-Based Reasoning in Finance: How Idealized Derivative Pricing Models Are Pragmatically Used to Evaluate Overall Market Efficiency**

This paper examines how no-arbitrage pricing models for forward contracts on non-dividend paying stocks and index futures are conceptually constructed and how they generate epistemically relevant knowledge about real financial markets. The analysis adopts a philosophical perspective grounded in Uskali Mäki's MISS account, which conceives models as idealized systems that isolate specific causal processes in order to function as credible surrogate representations of reality. The paper addresses two main questions. First, it investigates how no-arbitrage pricing models are structured through a derivation process that relies on strong idealizing assumptions, such as frictionless markets, uniform taxation, and perfectly efficient arbitrageurs. These assumptions are interpreted as isolating devices whose purpose is not empirical accuracy but the purification of a causal mechanism—here, the logic of arbitrage in efficient markets. Second, the paper explores how these idealized models are subsequently used in practice to provide guidance about real trading conditions and market behavior. By analyzing the standard derivation of forward prices on non-dividend paying stocks the aim is to show how arbitrage reasoning establishes a theoretical benchmark price that must hold if riskless profit opportunities are to be excluded. This benchmark functions as an ideal reference point rather than a direct description of actual market prices. Drawing on the notion of "credible narratives" as mediating devices, the paper argues that empirical data and historical events are used to pragmatically contextualize no-arbitrage index futures models. In this way, the models gain epistemic relevance not by accurately depicting markets but by highlighting, through contrast, the technological and operational factors that disrupt market efficiency. The paper suggests that no-arbitrage pricing models function both as

theoretical isolations and as practical tools for understanding real-world markets and trading facilities.

Chiffi Daniele; Giovanni Tuzet – **Abduction and the Precautionary Principle**

Decision-making in environmental and technological risk contexts is often shaped by deep scientific uncertainty and by the prospect of severe or irreversible harm. In such cases, the Precautionary Principle is frequently invoked to justify early interventions. Its application, however, depends crucially on how the plausibility of uncertain threats is assessed. This paper argues that abductive reasoning may offer a fruitful epistemic framework for grounding precautionary judgments in a principled and transparent manner. After briefly situating the Precautionary Principle within contemporary philosophical debates, the paper focuses on the epistemic task of evaluating uncertain future risks that are neither merely speculative nor fully established. Abductive reasoning is proposed as a method for assessing the plausibility of precaution-relevant hypotheses under conditions of limited evidence and high stakes. We investigate classical Peircean schemas of abduction and discusses their validity and possible limitations when applied to future-oriented and normatively charged contexts characteristic of precautionary reasoning. Particular attention is devoted to an adapted version of the Gabbay–Woods schema of abduction, which is argued to better capture the role of abductive reasoning in assessing potential threats and guiding preventive action. This framework allows abductive assessment to be responsive not only to epistemic considerations but also to precautionary aims, including risk mitigation and proportionality. The final part of the paper investigates how abductive evaluations of risk contribute to the rational assessment of precautionary interventions. Proportionality is analysed from both epistemological and legal perspectives, with specific attention to the role of cost–benefit considerations and balancing techniques in the justification of precautionary measures. The paper aims to clarify how abductive reasoning can strengthen the epistemic foundations of rational precaution under uncertainty.

Crupi Vincenzo – **Towards Synergistic Human-AI Interaction: A Bayesian Approach**

Can human and artificial intelligence effectively cooperate, so that their combination yields more accurate judgments than either alone? Research shows that this attractive outcome, labelled "Synergy" (Vaccaro, Almaatouq, and Malone 2024), is not easily attained at all, even when the independent accuracy of both humans and algorithms is significant. Here we present a simple Bayesian framework to model an instructive class of problems, and prove that a plausible and intepretable condition is mathematically necessary and sufficient for Synergy. This result provides a formal foundation to influential ideas in the literature and suggests that pursuit of Synergy, although more challenging than plain automation, is not vacuous. If sensibly connected with sound empirical investigation, our analysis also conveys additional insight to guide detection of the convenient antecedent conditions to aim at Synergy when possible and appropriate.

Danks David – **AI as Methods, Not (Only) Models**

Discussions of artificial intelligence (AI), whether in academia (including philosophy) or the public, typically focus on models: how we can learn them from data, what we infer or produce with them, the conditions in which they generalize appropriately, and so forth. In this paper, I argue that this focus has been largely misguided: when we are talking about AI research and development, we should focus on AI methods, rather than AI models. Of course, models are the product of methods, and we sometimes select a method precisely because it outputs certain kinds of models. Nonetheless, modern AI is best understood as primarily focused on developing novel (scientific) methods, rather than creating new models. In this paper, I first argue that most attempts to define 'AI' are systematically ambiguous about whether models or methods are the main focus. I then provide an empirical analysis of the papers published in major AI conferences, showing that methods are the focus in practice. Time permitting, I will also explore some examples from industry development and deployment, again arguing that

the key effort is on the methods, not the models. I conclude by exploring the implications of this shift in understanding: how should we change philosophy about AI, as well as public debates including those on regulation and governance?

Datteri Edoardo – **A Representational View of Biorobotics**

This paper argues that certain approaches to modelling animal behaviour using robots require representational and non-enactive methodological assumptions. The specific claims made here concern the validity of the surrogative stimulation strategy in interactive biorobotics. If this thesis is correct, it reduces the relevance of certain forms of robotics to the debate about embodied and enactive cognition, and suggests exercising caution when supporting these views about animal and human behaviour by relying on the success of robotics.

Dellantonio Sara; Luigi Pastore – **Model-Based Reasoning and Experimental Constructionism: The Operationist Legacy in Contemporary Psychology**

Model-based reasoning plays a central role in contemporary psychology and cognitive research by mediating between theory and data and enabling explanation across experimental contexts. Its success depends on the coordination of models, measurements, and phenomena. This paper argues that the enduring legacy of operationalism poses a significant challenge to this coordination. Although often dismissed as a failed epistemological doctrine, operationalism continues to shape psychological and cognitive research through task-specific operationalizations of constructs. Drawing on recent work that reconceptualizes operationalism as an exploratory epistemic practice, the paper examines how psychological research objects are constituted through clusters of experimentally generated phenomena rather than given as stable research object. While this approach facilitates exploration of poorly understood domains, it also leads to the proliferation and fragmentation of constructs tied to specific tasks and measurement procedures. From the perspective of model-based reasoning, this fragmentation threatens the taxonomic coherence required for general explanation and integration across models. By contrasting this operationist framework with realist accounts that rely on stable, cross-perspectival phenomena, the paper argues that empirical success in cognitive science may often reflect local coherence between models and measurements rather than robust tracking of underlying cognitive structures. The paper concludes by revisiting a classic dilemma concerning the trade-off between certainty and generality, suggesting that the intrinsic vagueness of psychological concepts should be understood as a constitutive feature rather than an epistemic defect.

Drago Antonino – **Reasoning by Means of an 'Adjunction': Its Relevance for the History of Mathematics, Logic and AI**

This communication is divided into three parts and an application to AI. In the first part, the intuitive notion of adjunction is recognized in several scientific and humanistic theories of great importance in the history of science and human culture. In particular, it is recognized in the history of infinitesimal analysis. Lazare Carnot attributed the genius of this astonishing calculus to what can be called an adjunction. In the second part, L. Carnot's methodological suggestions about the adjunction are translated into a reasoning method, represented graphically by a cycle. The notion of the reasoning cycle has been ignored by past scholars, who have attributed importance only to the deductive reasoning. This argumentation cycle is successfully compared to the inferential processes proposed by Charles Peirce; indeed, he substantially anticipated this novelty. The connections with 4E cognition theory of and the four scientific approaches to solve a problem are illustrated. The third part demonstrates that every adjunction is a doubly negated proposition whose corresponding affirmative proposition lacks evidence; therefore, it pertains to intuitionist logic where the double negation law fails. Through a comparison of the aforementioned theories, it is shown that, compared to axiomatic organization, an adjunction introduces a new way of organizing a scientific theory: the problem-oriented organization. Its theoretical development consists of four logical phases governed by intuitionist logic. As an application to AI it is noted that, since behavior is

governed by classical logic and includes at most statistical operations, a computer cannot implement doubly negation propositions and the new theoretical organization. This result logically characterizes in formal terms Peirce's philosophical suggestions about "impotences" of "thinking machines." It is therefore a recently formulated type of logic, intuitionist logic, which assures humanity's superiority over automated machines.

Fabbroni Claudio – **A Pragmatic Account of Neural Computations in Cognitive Models**

A central tenet of the computational theory of mind is that cognition is computational in the literal sense that neural systems implement computations to function. This talk challenges that assumption by arguing that computational models should not be interpreted in a realist, metaphysically committed way. Instead, I propose a pragmatic account according to which computational descriptions function as epistemic artefacts: idealized modeling tools that simplify the brain for intelligibility. Firstly, I argue that computational models do not uncover hidden computational structures inherent in neural systems, but rather construct selective patterns shaped by methodological and pragmatic choices. Marr's famous model of early vision exemplifies this: while capturing certain regularities in neural responses, it also omits empirically documented features that could have been treated as explanatorily relevant. Thus, the identification of a cognitive capacity with the computation of a specific function depends on selective abstraction rather than the discovery of an intrinsic structure. Moreover, even if neural systems are granted to compute, serious epistemic problems arise concerning the individuation of the computations they are said to implement. Multiple, equally adequate computational descriptions are compatible with the same empirical data, and realist theories of implementation face persistent difficulties in ruling out triviality arguments. Finally, I challenge the assumption that neural computation is medium-independent by highlighting the central role of chemical signalling in neural information processing. Molecular signaling depends essentially on the material properties of its vehicles and cannot plausibly be treated as medium-independent. Taken together, these considerations motivate a pragmatist account of computational models, which are thus seen as indispensable tools for prediction, understanding, and intervention, but should not be interpreted as revealing the essence of cognition. Rather, they function as idealized epistemic artefacts that manage complexity and support scientific understanding within model-based neuroscience. I claim that this approach is not only ontologically less committing but also closer to actual scientific practice.

Fabretti Annalisa; Alessandra Pelloni – **Mathematical Formalism, Narrative, and Normative Camouflage in Economic Modeling**

Economics represents a paradigmatic case of model-based reasoning, indeed it has historically placed highly idealized and abstract models at the core of theoretical analysis. While recent developments — most notably the empirical turn and the rise of behavioral economics — have diversified methodological practices, they have not displaced the privileged epistemic status of formal models. Influenced by broader shifts in the philosophy of science toward more naturalistic and practice-oriented approaches, contemporary economic methodology increasingly acknowledges methodological plurality, towards a large use of data and experiments, yet mathematical modeling remains a dominant marker of rigor and scientific legitimacy.

This paper stays at the intersection of epistemological analysis and normative critique, focusing on the role of narrative in mediating the epistemic authority of economic models. Drawing on the literature on model-based reasoning, we argue that models are not self-explanatory artifacts but require interpretive narratives that connect formal structures to economic meaning and policy relevance. Through such narratives, models function as mediating devices that construct "credible worlds", enabling explanation and guiding policy inference without purporting to offer literal descriptions of economic reality.

At the same time, the combination of mathematical formalism and narrative gives rise to a cluster of epistemic and normative risks. These include reification, whereby models are treated as direct representations of reality; methodological dominance, which marginalizes alternative approaches; and false precision, through which formal results convey unwarranted certainty

in open and historically contingent systems. These risks are inseparable from the value-ladenness of economic modeling. Values shape research questions, assumptions, standards of evaluation, and even the data on which models rely. Mathematical formalization can therefore function as a form of normative camouflage, embedding ethical and political judgments within technical choices that appear analytically necessary rather than contestable.

Rather than rejecting abstraction or formalization, this paper argues for greater reflexivity regarding the narrative-mediated and value-laden character of economic models. By linking their epistemic functions to their ethical and political implications, the paper advocates increased transparency and methodological pluralism in economic inquiry.

Fasoli Marco; Marco Viola; Agostino Pinna Pintor – **The More, the Merrier? Toward a Theory of Artifacts' Multifunctionality**

Function is a salient aspect of artifacts. Indeed, we would not have chairs and tables, forks and knives in our kitchens unless they did something for us. Theories of artifacts' function mostly disagree on the role of intentions in function determination (cf. Baker 2007, Evnine 2016, Houkes & Veermas 2010, Preston 1998). What they seem not to disagree on is the following assumption: however identified, artifacts normally possess a single proper function, the one for which they are designed or reproduced for. Chairs are designed for letting people sit, tables are there for holding stuff over, etc. Call such an assumption the proper mono-functionality dogma (PMD). The first aim of the paper is to challenge PMD, for PMD misrepresents the functional complexity of artifactual kinds due to perceptual and conceptual tendencies that make us myopic toward the functional complexity of artifacts. The second aim of this paper is to develop a systematic account of artifacts' multifunctionality. To date, philosophers working on the metaphysics of artifacts have largely neglected this aspect, which has not been considered a topic deserving systematic examination. However, once we abandon PMD, several questions concerning multifunctionality become immediately pressing. What does it mean for an artifact to be multifunctional? What is the relationship between multifunctionality and artifacts' affordances? Overall, our contribution aims at paving the way to a new, nuanced way to think about artifacts and their functions - emphasis on the plural.

References

Baker, L. R. (2007). The metaphysics of everyday life: An essay in practical realism. Cambridge: Cambridge.

Evnine, S. J. (2016). Making objects and events: A hylomorphic theory of artifacts, actions, and organisms. Oxford University Press.

Preston, B. (1998). Why is a wing like a spoon? A pluralist theory of function. The Journal of philosophy, 95(5), 215-254.

Feldbacher-Escamilla Christian J.; Alexander Gebharter; Michał Sikorski – **A Causal Theory of Suppositional Reasoning**

Suppositions can be classified as indicative vs. subjunctive and full vs. partial. We propose a causal account of suppositional reasoning that naturally unifies all four types of reasoning based on this classification, provides a justification of the rather heterogenous canonical update rules, and gives rise to a new update rule for the partial subjunctive case in terms of generalized imaging.

Ferrara Caterina; Iacopo Matteacci – **Algorithmic Partners: Epistemological Perspectives on Co-Created Knowledge in Human–AI Scientific Discovery**

The integration of artificial intelligence into scientific research is profoundly reshaping the processes through which knowledge is generated, validated, and interpreted. This paper examines the epistemological implications of Human–AI collaboration in scientific discovery, introducing the concept of Algorithmic Partners to describe AI systems as active contributors to knowledge construction rather than mere analytical tools. The central epistemic challenge lies in understanding how co-created knowledge between humans and AI systems reframes

traditional notions of explanation, understanding, and epistemic responsibility. The analysis adopts an interdisciplinary approach, drawing on philosophy of science, epistemology, cognitive science, and applied AI studies. Through a critical examination of published research and well-documented cases of Human–AI collaboration in data-intensive domains—such as computational genomics, complex-systems physics, and materials discovery—the paper highlights how AI contributes to hypothesis generation, the design of simulation-based experiments, and the identification of previously unseen patterns. This approach enables an examination of AI's epistemic role and its impact on expanding the cognitive capacities of human researchers, without the need to conduct original experiments. The paper underscores the epistemological opportunities afforded by algorithmic partners: accelerated discovery, the production of non-trivial insights, and enhanced robustness and replicability of scientific models. At the same time, it addresses risks and limitations, including algorithmic bias, the opacity of complex models, and the potential over-delegation of epistemic judgment. These dynamics raise concrete challenges for scientific responsibility and for defining criteria of trustworthiness and transparency. Finally, the paper reflects on the future implications of Human–AI collaboration, proposing an examination of emerging epistemic roles, shared responsibilities, and governance methodologies for co-created knowledge. This work offers a rigorous conceptual framework that balances philosophical abstraction with applied concreteness, providing an original, timely, and forward-looking perspective on the redefinition of scientific knowledge in the era of Human–AI collaboration.

Finkeissen Ekkehard – **Cuts, Quantization and Surrogate Models**

This contribution examines the often overlooked epistemic operation underlying all model-based reasoning: the setting of a cut. A cut is the decision to discretize a continuous process, to define boundaries, thresholds, or variables that do not exist as such prior to measurement. Cuts make quantization possible and generate the surrogate entities on which models operate.

I analyze cuts as hybrid epistemic actions with physical, conceptual, and abductive dimensions. They shape which hypotheses can be formulated, which simulations are meaningful, and which explanations become available.

The paper argues that understanding cuts is essential for evaluating the epistemic power and limitations of both scientific models and AI systems, since many pipelines implicitly hard-code such decisions without acknowledging their constructive role.

Fogarin Tobia; Matteo De Benedetto; Gustavo Cevolani – **Reconciling Epistemic Utility Theory and Truthlikeness: An Alternative to Proximity**

Bayesian epistemology studies norms governing degrees of belief. Epistemic utility theory provides a framework to justify some of these norms, such as probabilism, conditionalization, and the principal principle. The standard strategy in epistemic utility theory is to show that if a credence function violates a given norm, there exists an alternative credence function that satisfies the norm and is epistemically more valuable independently of what turns out to be true (Pettigrew 2016). Epistemic value is understood as depending solely on how well a doxastic state represents the world, irrespective of the consequences of acting on it. On this approach, all epistemically relevant considerations in a given context are captured by an epistemic utility measure, which is then required to satisfy certain general conditions. However, not all plausible sources of epistemic value appear to be compatible with the general conditions that epistemic utility measures should satisfy. A paradigmatic example concerns the conflict between truthlikeness and Strict Propriety.

The tension between truthlikeness and epistemic utility theory was first highlighted by Oddie 2019. Oddie proposes a condition on measures of epistemic value, Proximity, intended to capture considerations about truthlikeness, and argues that Proximity is incompatible with Strict Propriety, a condition assumed in most arguments in epistemic utility theory. Other authors have tried to solve this incompatibility by proposing weaker versions of Proximity (Schoenfield 2019, McCutcheon 2023). We propose an alternative way to account for the epistemic value of truthlikeness. We argue that our approach is more faithful to the truthlikeness literature than existing alternatives, and we show how to construct epistemic utility measures that value truthlikeness in our sense while also satisfying Strict Propriety.

Forghani Mohsen – **Contradictory Affordances in Local Context: A Topological Framework**

This paper develops a topological framework for contradictory affordances—situations in which an environment simultaneously invites and inhibits action within the same local setting. Paradigmatic cases include an emergency door marked EXIT that is chained shut, or a chair that ordinarily affords sitting but, when placed at the edge of a cliff, locally affords both sitting and avoidance. We argue that such cases are not primarily semantic paradoxes or belief-level inconsistencies. Instead, they instantiate practical contradiction: a conflict in the action landscape presented to an agent, where locally relevant invitations cannot be jointly honored without breakdown, hesitation, or revision.

Formally, we represent a scene as a finite set of environmental items and impose a neighborhood structure that captures local context—which elements count as relevant together for affordance evaluation. This neighborhood structure can be parameterized (e.g., by a spatial or task-relevance radius), allowing the framework to express how small contextual shifts can change what an object affords. Within a neighborhood, we treat affordances as action-guiding constraints rather than mere labels. A neighborhood exhibits contradictory affordances (relative to an agent and task framing) when the locally prioritized constraints cannot be jointly satisfied by any feasible course of action. This criterion distinguishes contradiction from ordinary affordance competition: competition concerns selecting among compatible, feasible invitations, whereas contradiction marks incompatibility that calls for revising the local model of action.

The framework provides two diagnostic refinements. First, minimal contradictory neighborhoods identify the smallest configurational subset responsible for conflict, avoiding trivial contradictions produced by overscoping. Second, varying neighborhood scope supports a robustness distinction between structural contradictions that persist across a range of contexts and fragile contradictions that appear only under narrow scoping. These resources connect affordance theory to topological tools for locality, stability, and emergence. We conclude by characterizing resolution strategies—rescoping context, reprioritizing constraints, shifting task framing, or intervening on the environment—and briefly noting how the account could later interface with computational detection without making empirical implementation a prerequisite for the conceptual contribution.

Galdiero Simone – **Modeling Reasoning Under Constraints: A Proof-Theoretic Approach to Contextuality and Hyperintensionality**

Classical logic models reasoning as globally invariant: propositions with the same truth conditions are freely interchangeable across all contexts of inference. While this idealization has been enormously successful, it fails to capture forms of reasoning in which context sensitivity, resource limitations, or fine-grained distinctions between propositions play a central epistemic role. This contribution explores how such constrained forms of reasoning can be modeled proof-theoretically, focusing on two prominent departures from classical logic: quantum logic and hyperintensional logic. Although quantum logic and hyperintensional logic originate in distinct domains (the foundations of physics and the philosophy of language, respectively) they both reject the inferential liberalism of classical logic. Quantum logic models reasoning under contextual constraints imposed by measurement settings, while hyperintensional logic preserves distinctions between propositions that are necessarily equivalent but differ in meaning, explanatory role, or inferential behavior. This work argues that these phenomena can be understood within a unified framework by treating them as consequences of systematic restrictions on admissible inference rules, rather than as effects of semantic enrichment.

The technical core of the contribution is a conservative extension of minimal quantum logic (LMQ) with a restricted implication connective inspired by the hyperintensional system HYPE. The resulting calculus preserves contextual sensitivity while exhibiting hyperintensional behavior. Therefore, even inferentially equivalent formulas are not freely substitutable under implication. Proof-theoretic results show that cut remains admissible and that classical inferential power is not reintroduced indirectly. From the perspective of model-based

reasoning, these proof-theoretic systems function as formal models of constrained reasoning practices. They allow systematic exploration of how limitations on inferential resources shape what can be inferred within a given context. While no claims of psychological realism are made, the framework is relevant to scientific, cognitive, and AI-related modeling, where reasoning is often context-dependent, tool-assisted, and structurally constrained.

Garavaglia Fabrizia Giulia; Marco Giunti – **Attention Heads and Pyramidal Columns: A Neuromorphic Model**

Is it possible to conceive of physical realizations of the processes that take place in the Transformer Module? More specifically, which neural structures could give rise to forms of processing of the kind carried out by Attention Heads? This paper addresses these two specific and fascinating questions. We proceed from the observation that the extraordinary verbal efficiency of large language models has something interesting to tell us about the processing of verbal information. It is fruitful, in our opinion, to establish specific connections between the functional modules of the Transformer and the functional areas of the nervous system that are involved in the proper execution of linguistic ability. Under this perspective, we highlight some morphological and operational correlations and we explore in more detail a new and bold hypothesis: The role played by an Attention Head in a layer of the Transformer Module, can be fulfilled by a pyramidal column of the Wernicke's area, one of the most important cortical parts involved in human verbal capacity. To deepen this hypothesis, we focus on analyzing the operations of the attention head matrices and the input projections they induce, and we dive into the functional organizations of the nervous system in search of elements of neural communication that can actually implement these processes. In the first part of our analysis, therefore, we elaborate a novel interpretation of the attention head as a system of neural networks with non-standard interconnections and then we show how this can be effectively mapped on the pyramidal columns of Wernicke's area. Our research highlights several features involved in neural communication within a pyramidal column and between different columns that support our hypothesis, showing how it is possible to recognize significant structural mappings between a Transformer and the nervous system in processing verbal information.

Gebharter Alexander; Barbara Osimani; Michał Sikorski; Elisa D'Olimpio; Zhitao Zhang – **A Causal Account for Estimating Effects Across Populations When No Causal Information Is Available**

Using experimental results obtained in a study population to estimate causal effects in a target population is key in science and evidence-based decision making. This task is especially challenging when the values of some variables are distributed differently among individuals in the two populations. We develop an account for effect estimation across populations that combines causal models and Jeffrey conditionalization and argue that in certain circumstances causal effects in the target population can be estimated without prior knowledge of the underlying causal structure. This is a major advantage, as obtaining causal information is often difficult, particularly in policy-making contexts.

Giordani Alessandro; Luca Mari – **Model-Based Inference in Measurement and AI**

Measuring systems and learning systems are increasingly treated as epistemically comparable sources of information about the world. This paper develops a systematic comparison between those kinds of devices by analyzing both as instances of model-based inferential systems, producing information based on structured inferences mediated by models, idealizations, and assumptions that connect observable data to target properties. The analysis begins with a structural account of measurement, according to which a measurement process involves the specification of a measurand, the construction of general and specific models linking the target property to empirically accessible quantities, and the explicit representation and propagation of measurement uncertainty. Within this framework, objectivity is understood as object-relatedness, and intersubjectivity as subject-independence secured by public models, standardized procedures, and metrological infrastructures. Learning systems are then

analyzed in parallel terms. The design of a learning system involves the selection of a hypothesis class, a loss function, and an optimization procedure, each of which embodies substantive modeling assumptions. While learning systems also rely on indirect inference, idealization, and uncertainty, their target properties are often operationally defined through labels or proxies whose relation to the intended property is weakly theorized or context-dependent. Furthermore, uncertainty is rarely treated as an explicit epistemic object, and learning systems typically lack the standardized infrastructures that secure intersubjectivity in measurement. On this basis, the paper argues that the epistemic privilege traditionally accorded to measurement does not derive from the absence of modeling or judgment, but from the structure and justification of the models that mediate inference, and from the institutional frameworks that support their validation. Since measurement admits degrees of trustworthiness, learning systems can in principle increase their epistemic reliability by strengthening the justification, transparency, and standardization of their modeling assumptions. Objectivity and intersubjectivity thus emerge as precise criteria for assessing the trustworthiness of AI-based inferences and for evaluating their role in the production of justified knowledge.

Giovagnoli Raffaela – **Perspectives on Habitual Behavior**

Habits have a very important function in individual life because they reduce the complexity of daily life; they make our daily life easier and more pleasant [1]. We choose habits regarding the satisfaction of our basic natural needs. Depending on the natural and social environment, we develop different habits that organize the way to satisfy our needs and desires [2]. We have habits in I-mode; for example, we practice meditation at sunrise or sunset or make a cake on New Year's Eve to bring good luck for the new year. We have habits in We-mode: a pleasant dinner with friends; the breather needs the collaboration between handler and dog before an exhibition.
It is essential to grasp a plausible notion of habitual behavior that can root the human reduction of complexity in ordinary practices. First, we need to overcome the so-called "Received View" [3]. According to the exponents of this vision, the only notion that explains the mechanism of habit is automaticity. In the words of Gardner [4] a psychological operationalization of habit has emerged, which incorporates an explanatory mechanism: habits are actions that are often performed because they have started automatically.

References

1. Giovagnoli, R.: Habitual Behavior: Reduction of Complexity of Human Daily Life. In Giovagnoli R., Lowe R. (Eds), The Logic of Social Practices II, Springer Sapere, Cham (2024)
2. Magnani L.: Ritual Artifacts as Symbolic Habits. In Giovagnoli R., Dodig-Crnkovic G. (Eds) Habits and Rituals Special Issue , Open Info Sci. De Gruyter, vol.3 115-126 (2018)
3. Douskos C.: Deliberation and Automaticity in Habitual Acts. Ethical Progress (1), 5-20 (2017)
4. Gardner, B.: Habit as automaticity, not frequency. Euro. Health Psycholog. 14, 32 (2012)

Gordillo García Alejandro – **Evolutionary Experiments, Islands, and the Epistemology of Reflective Scientific Models**

This paper examines a distinct kind of scientific model characterized by the fact that, besides representing their target system, also seems to instantiate it. I use experimental evolution, in which populations are propagated over thousands of generations in laboratory settings, as a paradigmatic case of these reflective models. Recently, philosophers like Marcel Weber have argued that these systems should be understood not merely as experiments, but as a form of experimental modeling, in which living organisms are in vivo representations of evolutionary processes. Building on this proposal, I argue that laboratory populations (e.g., E. coli) are not simulations but genuine evolving entities undergoing evolutionary processes. Their evolution is not more "artificial" or less "evolutionarily genuine" than evolution occurring in the wild. To support this claim, I draw upon the well-established framework of islands as model systems. Just as islands function as bounded, tractable sites where evolutionary processes occur in situ, evolutionary experiments are observable instantiations of evolution. This, however, creates

philosophical challenges: if evolutionary experiments are tokens of the type "evolutionary process," in what sense can they also represent that type? I explore two possible interpretations: first, that they enable second-order representation by generating abstract data (e.g., phylogenies, trait trajectories) from which general models are built; second, that they do not represent at all. Finally, by drawing general lessons for the epistemology of scientific models, this paper invites reflection by asking: where else in scientific practice does a model is an instantiation of its target system?

Graboń Wojciech; Marcin Woźny – **Model-Based Reasoning and Legal Interpretation: The Case of Rational Agents**

Legal practice routinely attributes rational agency to entities that are neither natural persons nor straightforward collective subjects, including legislators, courts, states, corporations, and paradigmatic figures such as the "reasonable person." These attributions structure legal interpretation, responsibility ascription, and institutional coordination. Yet they are rarely analyzed within a unified framework that connects social ontology, philosophy of models, and theories of legal reasoning.

This paper proposes such a framework by treating legal agents as model-based social entities constructed through practices of abstraction and idealization. Rather than asking whether these entities literally possess mental states, the paper adopts a modeling perspective according to which legal agents function as representational constructs whose rationality is specified by idealizing assumptions and functional parameters. In this respect, legal reasoning itself can be understood as a form of model-based reasoning: inferences are drawn not directly from social reality but from constructed representational artifacts that stabilize interpretation and enable surrogate reasoning in complex normative environments.

Building on idealized rationality conditions developed within the "humanistic interpretation" framework, the paper introduces a simplified three-regime typology of legal rationality. The first regime, Olympian rationality, models legal agents as fully informed and globally optimizing decision-makers and functions as a top-down interpretive idealization exemplified by figures such as the ideal legislator and Dworkin's Hercules. The second regime, bounded rationality, models real institutional actors as operating under informational and cognitive constraints and captures behavioral and institutional approaches to legal decision-making. The third regime, ecological rationality, reconceptualizes standards such as the "reasonable person" as context-sensitive heuristics optimized for coordination and norm compliance rather than ideal cognition.

The paper further distinguishes top-down and bottom-up modeling strategies and shows how legal institutions typically combine both, producing hybrid forms of agency. It argues that this integrated framework clarifies the methodological status of legal rationality attributions and extends model-based epistemology to institutional normativity, offering a unified vocabulary for comparing interpretive, institutional, and computational models of legal agency.

Grasso Leonardo – **From Algorithms to Architectures: Warranting Cross-Realization Generalization in Computational Cognitive Modeling**

Computational cognitive modeling is a paradigmatic form of model-based reasoning. We isolate a cognitive capacity, give it a computational characterization, and use a model to explain, predict, and generalize. Even without assuming that cognition is intrinsically computational, researchers routinely extend such models beyond their original targets and parameter regimes. This work asks what, methodologically, warrants cross-realization generalization. A common default, motivated by Marr-style levels of analysis and by the medium-independence of computation, treats implementation as deferrable. I argue that medium-independence for computational types does not entail implementation-independence for model-based inference. Explanatory and predictive success often depends on enabling conditions, including resource profiles, timing constraints, robustness requirements, stable primitives, and interface organization. These conditions are largely determined by computational architecture, which mediates between algorithm and substrate. The central proposal is an architecture-based criterion for warranted model extension: extrapolation

across realizations is justified only when an algorithmic model is paired with an explicit architectural surrogate that (i) identifies the organizational invariants relevant to the explanatory aim and (ii) provides principled grounds for expecting those invariants to be preserved across the intended target class. Making architecture explicit also constrains computational indeterminacy in physical systems by ruling out descriptions that presuppose unavailable resources or unrealistic cost structures.

Grasso Valeriano – **Context Drift and Normative Portability: How LLMs Reshape Epistemic Authority in Institutional Reasoning**

Recent discussions of LLMs in epistemology and AI research have focused on output reliability, bias, and misuse, treating epistemic risk as a function of model performance or user behaviour. I argue that such approaches overlook a distinct failure mode of model-based reasoning that emerges when language models are institutionalized as general-purpose epistemic. The epistemic risk associated with the institutional use of LLMs is a phenomenon I call context drift. It occurs when model-generated outputs, produced under specific epistemic and modelling assumptions, are stabilized and reused across heterogeneous institutional reasoning practices while being treated as context-invariant epistemic units. Crucially, not every form of reuse constitutes context drift. The phenomenon arises when a shift in the normative role of outputs — from exploratory/deliberative support to decision-guiding, justificatory, or regulatory content — is not accompanied by a reactivation of the assumptions that originally constrained their validity. From the perspective of model-based reasoning, this marks a transition from models as epistemic mediators to normative stand-ins within institutional settings. LLMs increasingly function as stabilizers of reasoning across contexts: their linguistically stable and general outputs are archived, circulated, and acted upon as if they retained epistemic legitimacy independently of the contexts in which they were generated. This enables a form of normative portability as a feature of institutional practices that confer epistemic authority on reusable model outputs without corresponding mechanisms of accountability. Context drift constitutes a limitation of model-based reasoning that cannot be captured by evaluations of representational adequacy, inferential power, or normative competence. Unlike misuse or anthropomorphic misunderstanding, it can occur even when models are understood and responsibly deployed. Introducing context drift as a constrained epistemological concept, I show that institutional embedding of language models reorganizes epistemic authority and responsibility, revealing why philosophical analysis is indispensable for assessing the limits of model-based reasoning in knowledge-producing institutions.

Gschwendtner Alexander – **Internal Representation and Inference: Model-Based Reasoning within Large Language Models**

Model-based reasoning is central to scientific practice: scientists routinely reason about target systems through surrogate models that stand in for those targets and support inferences about them. I argue that recent advances in Mechanistic Interpretability (MI) provide evidence that large language models (LLMs) likewise develop internal models, constructed by input data and tuning, that satisfy conditions for such surrogative reasoning. Drawing on teleosemantic accounts of representational content, I first establish that LLM internal states can possess genuine referential grounding, despite lacking direct sensorimotor contact with the real world. Training on human-generated text places these systems in mediated causal-informational relations to extra-linguistic reality, while preference-tuning constitutes a selection process that endows internal states with the function of tracking worldly conditions. I then examine mechanistic evidence that certain LLM internal representations exhibit three key characteristics of surrogate models: (i) standing-in for target domains, as demonstrated by entity-specific features that causally guide model behavior; (ii) structural correspondence, as shown by internal representations whose geometry mirrors the relational structures of target domains; (iii) support for counterfactual intervention, where manipulating internal representations produces systematically related changes in outputs. The heterogeneous architecture of LLMs, however, complicates straightforward attributions of model-based competence. Grounded mechanisms implementing principles execute concurrently with shallow heuristics, and outputs emerge from their aggregation. This calls for frameworks that

can characterize partial forms of model-based reasoning – i.e., attending not merely to whether systems possess such internal models, but to when and to what extent such internal models are actually operative in the generation of the output.

Guallart Nino – **It's Raining and You're Not Going to Die: Simulated Pragmatic Inferences in LLMs**

Large language models (LLMs) produce responses that, at the level of their linguistic role, simulate the incorporation of pragmatic inferences through which utterance underdeterminations are resolved by adding non-explicit content to the interpretation of the input. Although an LLM does not implement direct mechanisms to generate this kind of inference, training and alignment concentrate probability mass on pragmatically enriched outputs. We will discuss whether these role-level pragmatic inferences are better described from contextualist positions or from semantic minimalism, arguing that in most cases the former are more adequate. Building on the Recanati–Stanley debate between semantic minimalism and contextualism, we will argue that LLMs simulate enriching the meaning of the input in order to fix a useful reading in ways analogous to contextualist theses, rather than assigning contextual values to semantically articulated parameters and merely resolving ambiguities. In support of this claim, we explore the connection between these role-level pragmatic inferences and the actual mechanisms underlying text processing and generation, highlighting how self-attention implements a flexible context-integration mechanism that enables enriched readings of the input. We will also note that relevance theory can serve as a complement in explaining the selective pressure toward interpretations with greater communicative utility. Finally, we show how some hallucinations can be understood as a predictable effect of this pragmatization without world anchoring: when relevance and epistemic reliability cannot be simultaneously maximized, the system tends to favor over-enrichment and plausible contextual invention.

Harrell Maralee; David Danks – **A Model of (Conclusion) Confidence**

In this paper, we are concerned with developing a "logic" for confidence: a formal model that specifies the confidence that one ought to have in a conclusion, given a (possibly complex) argument—deductive or non-deductive—for that conclusion and confidences in each of the premises. We model confidence in a statement S as occurring in the [-1, +1] interval, where +1 denotes full confidence in S; -1 denotes full confidence in ¬S; and 0 denotes complete lack of confidence. We will refer to the value of our confidence in S as its "confidence-value." We consider our intuitions about how confidence in premises translates to confidence in a conclusion for some simple deductive and non-deductive arguments. These various observations point to a consistent formal model of confidence. We first develop that formal model under the assumption that the confidence-values of the premises are independent. We then relax that assumption in a generalization of the framework. Finally, we show how to further generalize the framework from "one-level" to "multi-level" arguments.

Hieke Willi – **On a Cognitive Consequence Relation**

The notion of a consequence relation is central to logical reasoning, as it formalizes logical consequence relative to a given set of sentences. Consequence relations have been extensively characterized and investigated in classical logics such as propositional logic and first-order logic, as well as in a variety of non-classical logics, including many-valued and non-monotonic logics. Every consequence relation induces a consequence operator on sets of sentences. Prominent properties of the classical consequence operator include monotonicity, deductive closure, and idempotence. Indeed, non-monotonic logics are so named because their consequence relations are not required to satisfy the monotonicity property characteristic of classical logic. This work aims to characterize a consequence relation that reflects aspects of human cognition. It consists of two parts. The first part presents an extensive literature review of attempts to incorporate findings from cognitive theory into formal logics, from both model-theoretic and proof-theoretic perspectives. This review also surveys empirical results from

behavioral experiments indicating systematic divergences between classical consequence relations and observed human reasoning patterns. The second part discusses candidate properties of a cognitive consequence relation, building on both the theoretical and empirical insights summarized in the first part. The overall goal of this work is to contribute to the development of modeling cognitive reasoning using methods from formal logic.

Hasnes Beninson Zvi – **Formalization Does Not Entail Neutrality: Evolutionary Game Theory and the Sociobiology Controversy**

Model makers often find themselves in the following situation: in order for the model to be able to represent the dynamics of the target system, they have to make certain assumptions that pertain to the system. However, these assumptions do not derive from the data, and alternative assumptions can also be used. Since a decision must be made, it derives from considerations such as theoretical commitments of the modeler. The central concern of my paper is that in those situations where the target system itself is inaccessible, and models become the central instruments to expolre it, the assumptions that a model embodies may come to seem inherent to the system. The use of models can thus reinforces and reifies implicit assumptions and biases embedded in the models. My paper takes the early attempt to formulate EGT as a case study and explains why the deployment of EGT was available only to advocates of the sociobiology program during the controversy over it. The assumptions formalized in EGT are compatible with the evolutionary picture to which sociobiology adheres; this evolutionary picture was the very thing being disputed. Thus, only the side advocating sociobiology was able to utilize EGT models. The relations between sociobiology and EGT demonstrate that certain modeling assumptions render models more appropriate for certain purposes than to others, while at the same time, the increased use of a model reinforces the users' trust in the assumptions the model is based on. To be clear, my paper does not aim to adjudicate between these sets of assumption. Instead, the point of my paper is to demonstrate that for each strand of EGT, an alternative set of assumptions was available. My hope is that the examination of the chosen assumptions will reveal the theoretical commitments that motivated the choice.

Hermkes Rico – **Epistemic Feelings and the Dynamics of Knowing and Not-Knowing in the Context of Critical Thinking**

The spread of misinformation and systematic information gaps pose significant challenges for societies, particularly in the context of 21st-Century Skills and Critical Thinking. This paper investigates the cognitive dynamics involved in critical thinking, focusing on how epistemic feelings shape processes of knowing and not-knowing. Drawing on Inferential Theory, the paper illustrates how individuals navigate incomplete or contradictory information through inferential cycles that involve abductive, deductive, and inductive reasoning. It examines their connection to epistemic feelings, which play a pivotal role in tacit inferences. The paper also addresses the methodological challenges of integrating feelings into inferential processes and advocates for a more process-oriented framework of Critical Thinking, providing valuable insights into how individuals confront the challenges of misinformation and missing information.

Ippoliti Emiliano – **The Unreasonable Effectiveness, Reasanable Effectiveness, and Eventual Ineffectiveness Of Mathematics On Wall Street**

The question of mathematics' effectiveness in describing and predicting natural and social phenomena remains philosophically contested. Existing views may be grouped into three broad positions: unreasonable effectiveness (Wigner 1960; Hamming 1980), reasonable effectiveness (Mac Lane 1990; Lakoff and Núñez 2000; Longo 2005), and reasonable ineffectiveness (Cellucci 2015, 2022). In this paper, I argue that this issue takes a distinctive form in the modelling and prediction of Wall Street, understood as a symbol of global finance, where epistemic tools can have ontological effects.

Because financial markets are performative and reflexive, mathematics does not merely represent them passively. It also helps shape market expectations, decisions, institutional practices, and ultimately market structures. From this perspective, mathematical effectiveness in finance follows a cyclical pattern. It may initially appear unreasonably effective, insofar as models formalize real statistical regularities, behavioural tendencies, and arbitrage mechanisms, even when they rely on contested assumptions. As these models become institutionalized and agents coordinate around them, mathematics becomes reasonably effective. Yet this effectiveness can later turn into reasonable ineffectiveness, when recursive behavioural responses generate systemic fragility.

I illustrate this cycle through the case of Value-at-Risk (VaR) during the 2008 global financial crisis. The widespread use of VaR produced a systemic underestimation of tail risks, grounded in assumptions that failed under stress: efficient markets, Gaussian return distributions, rational behaviour, and historical stationarity. Its initial appearance of precision masked deeper radical uncertainty, turning epistemic success into epistemic failure.

Jetli Priyedarshi – **Abduction and Analogical Reasoning in Plato's Euthyphro**

Euthyphro offers abundant applications of Peircean abduction and analogical reasoning. The Euthyphro provides roots of Peirce's accounts of abduction and analogy as two of four types of reasoning. Whereas abductions and analogies are found in the domain of ordinary discourse in the Euthyphro applied to the construction of definitions of terms like "piety"; in Peirce they emerge in schematic form, embedded in the core of scientific reasoning. We can move backwards from Peirce to Euthyphro and apply the more formal accounts to the context of ordinary discourse. Socrates abduces from Abduction 1, that if Euthyphro claims that his action of prosecuting his father is pious then he knows what piety is, to Abduction 2: that Euthyphro can provide a definition of 'piety'. Then with Socrates' pointing out the inadequacies of the definitions offered by Euthyphro a chain of abductions and definitions ensue: From Abduction 2 Euthyphro abduces by Abduction 3 to Definition 1: 'piety' is prosecuting a wrongdoer. This definition being too narrow leads Euthyphro to abduce from Abduction 2 by Abduction 5 to Definition 2: 'piety' is what is loved by the gods. The possibility of the same act being pious and impious at the same time, because gods disagree, leads Euthyphro to abduce from Abduction 2 by Abduction 6 to Definition 3: 'piety' is what is loved by all the gods. Socrates now points out a vicious circularity in this definition by using a masterful analogy: It seems that piety is piety because it is loved by the gods, just as a thing is carried because someone carries it, and something is loved by the gods because gods love it. But to treat 'loved by the gods' as a synonym of 'pious' leads to a vicious circularity. And the dialogue progresses with more abductions and more futility.

Larese Costanza; Hykel Hosni – **Polya's Fundamental Inductive Pattern for Medical Diagnostics**

In 'Mathematics and Plausible Reasoning' (1954), George Polya offered one of the earliest systematic accounts of plausible, non-deductive reasoning patterns, among which the fundamental inductive pattern (FIP) stands out for its simplicity and generality. Polya begins from a familiar situation: a conjecture A implies a consequence B, while neither is yet known to be true. If B is shown to be false, classical logic (via 'modus tollens') allows us to conclude definitively that A is false. But if B is found to be true, the result is different: A is not proven, yet it becomes more credible. The FIP thus captures a form of plausible rather than demonstrative reasoning: confirming a predicted consequence strengthens confidence in a hypothesis without establishing it as true. Despite its relevance to scientific reasoning, this pattern has received comparatively little attention within the logical community.

In this paper, we propose a formal reconstruction of the FIP by introducing a variant within the framework of adaptive logics, a family of systems specifically designed to model non-monotonic and dynamically revisable inference. Formally, our system, $AL^{r}_{P^\star}$ is built on the monotonic lower-limit logic $P^\star$, a first-order preferential logic with an object-language default connective, and incorporates a set of abnormalities and an adaptive strategy that regulate the defeasible application of the FIP. The behaviour of the system is illustrated through non-trivial examples from medical diagnosis, where typical clinical signs

increase the plausibility of a disease without ever entailing it, and where alternative explanations or incompatible background information may suspend or defeat the inference. In this way, we demonstrate that Polya's reasoning schemas can be integrated into a precise formal framework capable of representing central features of medical diagnosis.

Lee-Sursin Julian – **Towards an Explanatory Cognitive Model for Blackwell Informativeness**

When agents need information to make decisions under uncertainty, how do they determine which sources are most valuable? Blackwell's theory of statistical sufficiency (1951, 1953) offers one answer: an information structure is more informative if it cannot be derived from another by adding noise. But people don't actually reason this way. We ask questions, seek answers in particular orders, and face cognitive limits. To address this gap, we embed Blackwell informativeness within a cognitive framework based on mental models and question-directed inquiry. At a coarse-grained level, our framework recovers Blackwell monotonicity (in classical, idealized cases). At finer-grained levels, it predicts cases where agents arrive at different decisions depending on their path of inquiry, even when their information sources are identical by Blackwell's criterion. The result, we contend, is a more accurate account of how people actually evaluate and use information. More broadly, it may that cognitive science can be leveraged to address shortcomings of agent-modeling.

Lissack Michael – **The Orthogonality Principle and the Limits of Surrogative Reasoning in AI Systems: When Computational Models Cannot Substitute for Existential Understanding**

Large Language Models function as computational surrogates for human reasoning, generating outputs that mimic the products of model-based cognition. Yet their capacity for genuine model-based understanding is fundamentally constrained by what this paper terms the "orthogonality principle"—the recognition that computational and existential dimensions of cognition operate along genuinely independent axes. Drawing on Hans Vaihinger's philosophy of "as if" and Robert Rosen's theory of anticipatory systems, we argue that AI systems cannot validate their models through embodied intervention in causal environments, rendering them useful fictions that require human existential grounding to function as genuine reasoning surrogates.

The orthogonality principle establishes that computational sophistication provides no information about existential understanding, ethical sensitivity, or epistemological awareness. Just as movement along the x-axis tells us nothing about position on the y-axis, pattern recognition capability tells us nothing about lived comprehension. These dimensions are non-translatable: we cannot transform computational pattern-matching into existential understanding any more than we can transform syntactic manipulation into semantic comprehension.

This framework explains why LLMs generate syntactically perfect but semantically problematic outputs: they optimize for statistical patterns rather than intervention success in causal environments. From the anticipatory systems perspective, LLMs are not anticipatory agents in Rosen's sense because they lack feedback loops that validate or refute their outputs through consequences in causally bounded environments.

The paper contributes to model-based reasoning scholarship by providing diagnostic criteria for distinguishing when AI systems function as useful fictions within human anticipatory systems from when they fail, and by clarifying the ontological status of AI-generated models as computational surrogates that require human existential completion. This analysis advances debates about models as fictions while offering principled criteria for distinguishing genuine surrogative reasoning from computational mimicry.

Lizzadri Antonio – **Extended Minds Changing Mind: Rationality and Identity Across Conceptual Revolutions**

This paper aims to provide an externalist explanation of the transformation of rational acceptability criteria and subjective identity across conceptual revolutions. I maintain that extending rational justification processes as well as the mind through the external environment allows us to better address some theoretical challenges posed by conceptual revolutions, that remain problematic within a mere internalist account. More precisely, I first argue that considering rational acceptability criteria as extended model-based dynamic cognitive tools instead of fixed intrapsychic epistemic norms it is possible to face the problem of arbitrariness of knowledge in paradigm shifts. In fact, if considering only intra-theoretical elements, alternative conceptual schemas could appear arbitrary since they involve different and ultimately incommensurable representations of the world, considering also extra-theoretical elements, conceptual schemas do not seem unconstrained since the transformation of the rational acceptability criteria is subject to constraints arising from interaction with the environment. Based on this result, I thus suggest extending not only cognition but also the mind in order to address a further theoretical challenge concerning the persistence of subjective identity. Again, while an internalist conception of subjective identity seems to have to resign to a loss of identity due to the radical change of internal mental states across conceptual revolutions, I suggest that the transformation of an extended identity can be understood as a gradual and intelligible transition from an extended cognitive ecology to another, through which the identity is secured not by the continuity of internal mental states that unavoidably change, but by the practice of self-adjustment to a dynamic environment, that the cognitive agents are able to retrospectively trace maintaining a properly processual self-awareness. In both cases, I maintain that a consistent hybridization of rationality and identity should acknowledge the intentional interdependence between the mind and the world.

Martínez-Cazalla Maria Dolors – **Model-Based Dialogical Reasoning with Refined Conjunctions: *With* and *But* as Cross-World Operators**

This paper develops a model-based extension of Dialogical Logic within the Rahman framework, grounding argumentative reasoning in a pragmatically constituted proof-theoretic semantics and a formal epistemology of interactive justification. Validity is conceived as the existence of winning strategies in regulated dialogue games.

The paper extends the standard framework of Dialogical Logic through the introduction of refined conjunctive operators, namely *with* and *but*. These operators provide a more precise formal tool for dialogue structuring and argumentative construction, enhancing clarity and strategic effectiveness in complex negotiation contexts. This technical development enables improved analysis of dialogue dynamics and supports the construction of stronger argumentative structures aimed at bridging different legal systems, and thus different worlds.

Building on previous work applying Dialogical Logic to precedents for change in the final legal sentence, the research demonstrates the potential to modify jurisprudential outcomes across Civil Law, Common Law, Islamic Law, Socialist Law, and Talmudic Law traditions. By identifying dialogical common and non-common denominators among these systems, the framework seeks to overcome legal incompatibility in international negotiation processes in both public and private law. The refined conjunctive forms play a central role in modelling coordinated and contrastive inferential relations across normative worlds; they can overcome non-common denominators, the main hurdle we have in international negotiation process.

More broadly, these new operators open perspectives for formal and computational reasoning, including applications in AI and emerging paradigms such as quantum programming, while strengthening the epistemic foundations of dialogical model-based reasoning.

Mela Quílez Marc – **When Deep Neural Networks Become Scientific Models. A Pragmatic Representational Approach to AI-Based Atmospheric Modelling**

The increasing use of deep neural networks (DNNs) in scientific modelling is transforming key epistemic functions traditionally attributed to models, such as prediction, explanation, and representation. Consequently, a growing debate has emerged about whether, and in what sense, DNNs qualify as “scientific models”. Some argue that, because of their opacity, DNNs lack the representational capacity required for scientific modelling and should instead be understood as optimisation tools. Artifactualist approaches reject representation as a

necessary condition for modelhood and explain the epistemic value of DNNs in terms of their design and use. I argue, however, that abandoning representationalism obscures how AI-based modelling supports inferences about the world. Therefore, I examine attempts to preserve a representational interpretation of DNNs, either by treating them as minimal input–output models or by drawing analogies with idealised models when combined with explainable AI techniques. While illuminating, these proposals generate systematic ambiguities concerning what counts as the model, what grounds scientific understanding, and which target is being represented. I show that these ambiguities arise from different ways of conceptualising DNNs—as complete functions, as input–output mappings, or as targets of explanation—and that these conceptualisations correspond to distinct epistemic roles. To resolve this tension, I advance a pragmatic yet non-deflationary account of scientific representation, on which representation depends on how scientists specify how model features are taken to stand for target features, given the context and purposes. I develop this framework through three case studies in atmospheric sciences, showing that DNNs do not fit into a single epistemic category. In some contexts, their use is instrumental and non-representational, while in others it supports genuine representational models. The pragmatic–representational framework proposed here provides principled criteria for distinguishing these uses and for explaining when—and in what sense—DNNs become scientific models in contemporary research practice.

Michelini Matteo; Dunja Šešelja – **A Fit-for-Purpose Epistemology of Highly Idealized Models**

Highly idealized computational models (HIMs), such as agent-based and other toy models, play a central role in contemporary social science and humanities. Despite their widespread use, there remains disagreement about how such models should be evaluated and what epistemic functions they serve, especially given their deliberate abstraction from real-world systems. This paper develops an inferentialist framework for evaluating the epistemic contribution of HIMs by treating them as argumentative devices whose function depends on the argument in which they are embedded.

We argue that the epistemic value of a highly idealized model cannot be assessed in isolation, but only relative to the specific claims it is intended to support. Accordingly, we propose evaluating models in terms of their adequacy to an argument: a model is fit-for-purpose insofar as it provides sufficient warrant for the conclusion of that argument. We adapt Parker's evaluative framework for climate models to the context of social-scientific HIMs, shifting the notion of adequacy from general empirical fit to inferential and persuasive strength within a given argumentative context.

We contrast this approach with accuracy-first views, which assess models primarily by their representational accuracy with respect to an actual, real-world target system. We argue that such views struggle to accommodate models that target non-actual systems, depend on context-sensitive interpretations of idealization, and obscure the fact that the same model may serve different epistemic functions in different arguments. Using Schelling's segregation model as an illustrative case, we show how a single model can support theoretical exploration or how-possibly explanation depending on its argumentative embedding. Our framework offers a unified, context-sensitive account of model evaluation that better captures the diversity of epistemic roles played by highly idealized models in social science and humanities.

Minnameier Gerhard – **Abduction And Abstraction – Peirce, Piaget and a Post-Kohlbergian Theory of Moral Development**

Peirce's pragmatism rests on the assumption that all knowledge is built on experience and is derived from this ground rather than from metaphysical principles or inborn categories. It starts with the early behavioural habits humans form in the interaction with their environment (e.g., CP 5.181, 1903), continues with the acquisition of self-consciousness and language and extends into the lofty hights of philosophical and scientific reasoning. In a similar way, although, as it seems, independently, Piaget developed his theory of the equilibration of cognitive structures, which is also a pragmatist approach (or a constructive one, as cognitive psychologists would prefer to say) inasmuch as it explains cognitive development in terms of the construction of knowledge based on experience. As will be shown, both approaches are

congenial in the sense that Peirce delivered a true developmental logic that accounts for stage transitions via abduction and Piaget set forth suitable principles to account for hierarchies of knowledge, in particular in his mature work (which is often overlooked). It will be shown how the insights from both sides can be used to explain moral development in the context of a post-Kohlbergian theory of moral stages and development.

Niño Douglas – **When Abduction Preserves Ignorance: On the Missing Upgrade Step in John Woods' Causal-Response Model of Knowledge**

John Woods' Causal-Response Model (CRM) reconceives knowledge as well-produced true belief: true contents are known when they are causally generated by information-processing under the right conditions, with properly working cognitive devices, good information, and no incapacitating negative externalities. At the same time, Woods maintains (in a Peircean spirit) that abduction is ignorance-preserving: abductive inference yields a conjecture or a reason to suspect a hypothesis, but it does not authorize belief. This generates a tension that has not been fully stabilized: if knowledge requires belief, and abduction preserves ignorance by withholding belief, then even perfectly truth-tracking abductive success seems unable to yield knowledge as such. Conversely, if abductive suspicion is treated as belief, the ignorance-preservation thesis risks collapsing into a weak-belief position.

This presentation develops two challenges for the CRM. First, it is necessary to clarify what "ignorance" means in the ignorance-preservation thesis: is it merely lack of knowledge, or lack of endorsement, commitment, methodological security, or intersubjective stabilization? Second, the transition from conjecture to belief is not the same in individual and collective contexts: scientific belief-formation is constrained by communal standards, distributed resources, and method-regulated uptake.

To address both issues, I propose an Agentive Semiotics approach that extends Woods' agent–agenda framework into an enactive-semiotic model of inference. The key move is to distinguish endorsement from commitment. Abduction produces graded recommendation and investigandum priority without licensing commitment; induction (in the Peircean, methodologically constrained sense) is a central route by which endorsement may cross into commitment. In collective contexts, commitment thresholds are intersubjective and depend on coordinated standards and shared resources. The result is an adjusted CRM: knowledge requires not only causal belief-production but an explicit architecture of stance transitions from conjecture to endorsement to commitment, and from individual upgrade to intersubjective stabilization.

Obshanska Vira; Mariusz Urbański; Marta Skorodziievska; Kateryna Shyiko; Aleksandra Ulaszewska – **The Influence of Stimulus Presentation Mode on Creativity Measures: A Modified Alternative Uses Test Study**

This study investigates how stimulus presentation mode influences performance on a modified Alternative Uses Test (AUT). Drawing on model-based reasoning frameworks that emphasise how representational formats shape cognition, we tested whether presenting objects as (1) name only, (2) photograph of the whole object with name, or (3) photograph disassembled into parts with name activates different cognitive mechanisms. We hypothesised that the disassembled condition facilitates 'perceptual decomposition' – disrupting dominant functional associations and enabling novel recombinations of components.

In our pilot study (N = 31), participants generated alternative uses for everyday objects (e.g., hammer, bottle, coffee maker) across the three presentation modes. We assessed fluency (number of uses), flexibility (number of categories), originality (judge ratings on a 1–3 scale), and uniqueness (proportion of non-repeated responses).

Results revealed no significant fluency differences between modes. However, the disassembled-parts condition yielded significantly higher originality than name-only or whole-object conditions—an effect robust to controlling for fluency. Uniqueness did not vary by presentation mode but decreased with higher fluency, revealing a tension between idea quantity and uniqueness. Category composition differed both between objects and across presentation modes, indicating that stimulus properties interact with presentation format to shape the solution space.

These findings suggest that perceptual decomposition scaffolds divergent thinking by emphasising structural components over functional wholes, thereby facilitating movement beyond typical object functions without inflating response quantity. The results have implications for understanding how external representations shape creative cognition. We are currently conducting an extended replication with 150 participants to test robustness across a larger sample and explore cognitive style as a potential moderator. Results from this larger study will be presented at the conference.

Osta-Vélez Matías – **The Price of Compression: Why Efficient Concepts Lead to Biased Reasoning**

Everyday reasoning frequently deviates from normative rules of logic and probability. While cognitive psychology has extensively documented these biases, existing models—such as dual-system theory or Bayesian approaches—often fail to explain the specific representational mechanisms behind them. In this talk, I argue that reasoning biases are not merely contingent errors, but structural consequences of the way semantic memory organizes and compresses information. I propose that concepts function as dimensionality-reduction structures, analogous to Principal Component Analysis (PCA), which allow the cognitive system to efficiently simplify environmental complexity.

Drawing on the theory of Conceptual Spaces, I propose to see concepts as regions where property covariation is exploited to identify latent axes (principal components). I argue that in contexts of explicit-fast judgment, the cognitive system prioritizes "structural coherence"—the alignment of exemplars with these latent components—over extensional criteria (like probability or logic). Biases thus emerge when the criterion of "good compression" conflicts with the rules of probability.

This framework is applied to two classic phenomena: the conjunction fallacy and inclusion fallacies. Regarding the conjunction fallacy, I show that a conjunctive concept (e.g., "feminist bank teller") often projects closer to the principal component induced by a description than a single concept ("bank teller"). Geometrically, the conjunction yields a representation that is more coherent and better compressed, despite being less probable. Similarly, regarding inclusion fallacies, the system prefers specific subcategories over superordinates because the former constitute more compact, informationally dense clusters.

Park Woosuk – **Magnani's Manipulative Abduction, Revisited**

Magnani’s Manipulative Abduction, Revisited Woosuk Park (KAIST) woosukpark@kaist.ac.kr (Abstract) There is no doubt that Lorenzo Magnani is the leading scholar in the field of abduction. Not to mention his monographs and articles on abduction, his dedication to the research community around abduction is awesome. Among his many contributions, however, the notion of manipulative abduction is particularly distinguished. By highlighting the importance of manipulative abduction in science and everyday life, Magnani not only goes with Peirce but also beyond Peirce. It is not merely an accident that manipulative abduction is a focal issue for anyone who discusses Magnani’s views on abduction. Time ripens to ascertain what exactly philosophers find interesting and significant in Magnani’s views on abduction. Some want to clarify and deepen the concept of manipulative abduction. Others find promising prospects of adopting Magnani’s views to enhance their own closely related concepts or theories. Still others even want to point out that manipulativeness is essential to abduction. The modest aim of this paper is to present a literature survey of all these invaluable attempts. In section 1, I discuss the context in which Magnani introduced manipulative abduction and identify some possible antecedents that inspired him. For this purpose, we note “the practice turn” in the history and philosophy of science, which includes the movement of emphasizing the importance of model-based reasoning in scientific practice. Section 2 will present a brief overview of how Magnani’s multiple distinctions to classify abduction are appreciated by others. In section 3, I will review discussions of manipulative abduction in diagrammatic reasoning in geometry. In section 4, given the success of the wide applicability of manipulative abduction, I will discuss the essential manipulativeness of abduction. I will conclude, in section 5, by indicating where and in what ways we may fruitfully apply Magnani’s manipulative abduction that already proved its mettle.

Pietarinen Ahti; Zhiyi Ye – **Triadic Information and Quasi-Minds: Model-Based Proto-Cognition in Aneural Xenobotic Collectives**

Are brain-like integrated information patterns specific to nervous tissue? We present a model-based reasoning account of semiotic emergence in aneural xenobot organoids. Building on recent analyses in philosophy of bioengineering (Pietarinen et al., 2025a,b) and calcium-signaling collectives (Varley et al., 2025), we assess the relationships of higher-order information architectures in xenobots to resting-state human brain networks. Multivariate information-theoretic models with stringent nulls that preserve local dynamics to break inter-element coordination appears to evidence non-trivial (triadic) dependencies, modular organisation, spatially coherent interactions, and alternating integration/segregation over time. These whole-level properties increase the system's predictive power beyond the sum of isolated parts, indicating coordination with emergent higher-order organisation.

We interpret these findings within a Peircean semiotic and logical framework: calcium patterns as signs, environmental/physiological constraints as objects, and collective tissue responses as interpretants. The presence of synergy (novel joint information) and redundancy (robust repetition) signals triadic semiosis and supports graded, substrate-independent accounts of proto-cognitive competence. This helps articulate an abductive hypothesis: "brain-like" informational organisation can arise in aneural living substrates to underwrite adaptive behavior without committing to panpsychism. Instead, we defend a synechist, continuity-based view of quasi-minds; continuity of organisation helps to avoid overextending "mind" and avoid combination problems (Beni & Pietarinen 2026).

Our current contribution is threefold. First, we show how surrogative modeling and null-model testing ground semiotic claims in quantifiable structures, linking semiotics to model-based explanation. Second, we propose experimentally tractable predictions arising from Varley et al. (2025) (e.g., gap-junction perturbations; spatial re-embedding; pharmacological modulation) and contrast them conceptually with expected null outcomes. Third, we outline a comparative test program for artificial systems (e.g., large language/reasoning models) to assess whether similar higher-order informational organisation emerges in non-living substrates. We conclude with a model-based taxonomy of emergent competences in xenobots: coordination, habit-formation, and symbolic-like coding in bioelectric signaling, and discuss implications for embodied cognition, organoid bioengineering, and AI benchmarks of Peircean "quasi-mind" type organisations.

Pinna Simone; Marco Giunti – **Partially Inferential Behavior and Cognitive Integration in Large Language Models**

Large Language Models (LLMs) often produce outputs that look like reasoning. Yet their performance remains fragile. Small changes in wording can disrupt results, constraints can be lost, and fluent arguments can be produced with little support. This paper develops a model-based and system-level explanation of why LLM outputs can appear inferential while remaining unreliable under constraint. The central proposal is that these outputs are best described as partially inferential behavior: across a multi-step continuation, the system keeps prompt-introduced constraints in play and uses them to guide later steps. This notion is contrasted with full inferential behavior, which requires control beyond local constraint maintenance, including sensitivity to validity and counterexamples, reliable evidence management, calibration, and accountability within a practice. The empirical discussion contrasts broad benchmark testing with constraint-tracking diagnostics. Benchmark suites enable large-scale comparison, but accuracy can be inflated by overlap, artifacts, or shortcuts and does not show whether task requirements are sustained across steps. Constraint-tracking diagnostics instead probe stability under controlled variation. ARC and ConceptARC exemplify this strategy and reveal a dissociation between rule articulation and rule execution. We then introduce a model-based account of how context-conditioned dynamics in decoder-only Transformers can support a model's ability to keep task constraints in play across several steps. At inference time, the prompt shapes relevant relations, and self-attention computes runtime variables that depend on the specific context. When requirements are stable enough, these dynamics can preserve dependencies across steps. Finally, the paper clarifies how partially

inferential behavior can contribute to full inferential practice when LLMs are embedded in distributed cognitive systems, understood as coordinated activity across people and representational artifacts. On this view, inferential behavior depends on how the coupled system organizes verification, control, and accountability over time, rather than on the model alone.

Pregel Fabian; Johannes Himmelreich – **Does AI Deception Risk Establish the Necessity of Mechanistic Interpretability?**

Mechanistic interpretability (MI) is often presented as 'necessary' for AI safety: without insight into a model's internal mechanisms, evaluators allegedly cannot rule out strategic deception that passes behavioural tests but fails in deployment. This paper evaluates that necessity claim—MINAS —by analysing the deception worry. We argue that deception risk does not establish MINAS ; at most, it supports the weaker conclusion that MI may be an important component of a layered assurance strategy. We first clarify the target phenomenon. Standard philosophical accounts treat deception as intention-involving; since it is contested whether present-day large language models have the relevant intentions or beliefs, talk of 'AI deception' is, strictly, controversial. Yet the underlying safety concern persists: systems may exhibit strategic misleadingness, i.e. verbal or non-verbal behaviour that leads evaluators to overestimate safety or alignment. Once framed this way, behavioural evaluation can indeed be gamed, but the inference to MI-necessity fails. Much misleadingness is, in principle, externally checkable through audits and independent verification. More difficult cases also do not yield necessity. Misleadingness about internal states (what a system 'knows', whether it is concealing information, why it acted) is not directly verifiable, but black-box methods—red teaming, cross-context consistency tests, and incentive-shifted evaluations—can supply indirect evidence and justify robust deployment constraints, analogously to familiar forms of human safety governance. Likewise, a system can mislead while stating only truths, via implicature, selective disclosure, and framing; nevertheless, structured cross-examination can force implicit commitments into the open and expose incoherence over time. Finally, two pragmatic complications reinforce the conceptual result: current MI techniques remain partial, and interpretability 'tests' can themselves become targets for gaming. Overall, the deception worry is best understood as an argument for layered defences, not for MINAS.

Prätor Klaus – **Two models of Artificial Intelligence**

Artificial Intelligence can be conceived and designs according to (at least) two different models. The nowadays prevalent and widely accepted is that of simulating human intelligence But there is and was always the alternative model of conceiving it as a human tool. You can find models of external intelligence long befrei the era of computers. And from the beginning of digital AI.to recent publications existed the alternative approach to model AI aa a human tool. We demonstrate it using the work of Terry Winograd - and some examples of systems, which can be modeled as a humanlike intelligence or as a tool - both with equivalent functionality. Especially dialogue systems or chats are easily perceived as intelligent. AI-critic Joseph Weizenbaum showed this in an early experiment. The first generation of AI was based on recursive list handling (cf. LISP LISt Processing) and especially logic inference (PROgrammation LOGique). Logic inference is only based on the logical form of the used rules without any understanding of their content. This is also the reason why it can be used in a mathematical (algebraic) way. The second generation of artificial intelligence using neural networks reinforces the impression of humanlike intelligence. But these systems depend on the exploitation of huge amounts of human generated texts. The function of AI is essentially a very refined and elaborated search tool ,which produces impressive results - but without humanlike understanding of language and facts.

Roldán Roldán Francisco Isidro; Matthieu Fontaine – **The Fauna of Rigidities**

This paper addresses the lack of systematicity in the concept of rigid designation, proposing a unified taxonomy to categorize the diverse and often conflicting "fauna" of notions present in

current literature. While the intuitive core of rigidity is widely accepted, its application across different modal frameworks often lacks the rigor necessary for coherent semantic definitions. The study pursues two primary objectives: first, to establish a precise classification of rigidity types based on their formal commitments; and second, to analyze how these distinctions influence key debates in modal logic and the philosophy of language. Methodologically, the proposed taxonomy organizes the "fauna" along three fundamental axes: 1. Domain Type: Distinguishing between constant-domain and varying-domain models. 2. Strength: Evaluating the requirements for weak rigidity (dependent on existence or the partiality of the interpretation function) versus strong rigidity. 3. Scope: Contrasting local rigidity (restricted to accessible worlds) with global rigidity (applicable to the entire model). The findings suggest that many persistent disputes between direct reference and descriptivist theories are rooted in the implicit adoption of non-equivalent notions of rigidity. By providing a systematic framework, this work clarifies the impact of these distinctions on formal principles such as the necessity of identity and existential generalization. Finally, the paper proposes that this classificatory model can be extended to the analysis of general terms, specifically natural-kind terms, providing a foundation for future research into the referential behaviour of non-singular expressions.

Ruyant Quentin – **Scientific Models are Not Logical Models**

The semantic conception of theories claims that theoretical models are truth-makers for theoretical statements: theoretical laws do not describe the world directly, but instead a family of non-linguistic structures that can later be compared with real systems. This is usually analysed using model-theoretic tools. I believe that this conception of theoretical models as truth-makers is confused. It stems from an ambiguity between two meanings of "model", as what represents and as what is represented. The fact that scientist usage doesn't match that of logicians in this respect was acknowledged, but dismissed by earlier proponents of the semantic view, thus introducing an inherent tension within the view. Attempts to resolve the tension by claiming that models can fullfil both roles depending on the activity (theoretical or experimental) results in an implausible and rigidly nested picture of scientific representation: these two kinds of activities are best understood as more or less abstract or hypothetical representations of the same world, which is compatible with a linguistic format all the way through. Finally, claiming that scientific representation is non-linguistic because it is either informal, pictorial, idealised or contextual, doesn't work either: structural representation is either useless or fares worse than linguistic representation in these respects. In conclusion, I examine how paying attention to pragmatics in philosophy of language provides better prospects for analysing the contextual aspects of theoretical representation in science.

Sans Pinillos Alger – **Managing Uncertainty in VUCA Environments: Wargames as Model-Based Reasoning Tools**

Decision-making in security and defense contexts is increasingly taking place in environments characterized by volatility, uncertainty, complexity, and ambiguity (VUCA). In these scenarios, traditional modeling approaches, geared toward prediction, optimization, and risk reduction, exhibit structural limitations that are exacerbated by the incorporation of decision-support systems based on non-deterministic and adaptive AI. Given that the behavior of these systems cannot be fully anticipated or exhaustively verified ex ante, the central epistemological challenge is no longer to eliminate ignorance through better data or more accurate models, but to manage uncertainty under conditions in which structural ignorance persists. This work argues that, in VUCA environments, model-based reasoning (MBR) should be understood as a non-predictive, exploratory, and normatively oriented practice. Rather than aspiring to reliable predictions about future states of the world, MBR functions to structure uncertainty, delineate plausible courses of action, and support critical decisions under incomplete information and time pressure. From this perspective, wargames play a central role. Despite their use in predictive simulations or training exercises, wargames serve as MBR tools that enable abductive reasoning about uncertain futures, the evaluation of decision thresholds, and the exploration of ethically and strategically relevant modes of failure. This thesis is articulated through three main contributions: a precise characterization of the relationship between

uncertainty and radical ignorance in VUCA environments; an abductive conception of reasoning grounded in the eco-cognitive model; and an analysis of wargames as second-order exploratory models that render decision architectures visible. The paper also introduces the role of bad expectations as negative heuristics that constrain the space of plausible actions and direct attention toward escalation risks and critical failure modes. Finally, it argues that wargames are fundamental epistemic testbeds for evaluating the robustness of decision patterns when traditional verification and validation approaches prove insufficient in VUCA environments with high strategic impact.

Schwartz Nora Alejandrina – **Studies of Innovative Scientific Reasoning and Situated Cognition**

Epistemology, understood as a discipline referring to scientific practices, has thematized the environment and the material artifacts that comprise them with different approaches. Within this panorama of epistemological approaches, I consider that certain cognitive studies of science of the "Kuhnian tradition" occupy a preeminent place. The reason for this is that they have specifically investigated practices creative of new concepts and hypotheses. These studies have been characterized by addressing problems and themes raised by Thomas Kuhn. One of the central issues Kuhn addressed concerns scientific revolutions and the conceptual change they entail. To develop this undertaking, they employed some of the categories of the situated cognition, an "environmental" cognitive sciences perspective. A large part of Nancy Nersessian's work belongs to the epistemological orientation just described. In this paper, I refer specifically to her work and emphasize that she investigated a class of practices, model-based reasoning, employing conceptual tools from a version of the situated cognition, the distributed cognition–although not exclusively. I emphasize that, in my opinion, this approach has allowed for an articulated conceptualization of the way in which "external" material representations shape scientific modeling. First, I will review Nersessian's approach to model-based reasoning. I will analyze how, with the aid of the perspective of the situated cognition, she has considered this kind of reasoning as practices for solving representational problems. Then, I will highlight that, relying particularly on the distributed cognition, she has conceived of modeling as an emergent property of the coordination of cognitive processes within a cognitive system. Finally, I will evaluate the extent to which Nersessian's conception of model-based reasoning contributes to clarifying the impact of material resources on cognition.

Segarra Adrià – **Models, Facts and Bayesian Confirmation**

John Norton has been one of the most influential critics of Bayesianism in contemporary philosophy of induction, advancing his Material Theory of Induction (MTI) as an alternative that grounds inductive support in domain-specific facts rather than universal formal rules. This article shows that the apparent opposition between Norton's material approach and Bayesian inductive logics is misleading. I argue that, once Bayesianism is properly framed within Segarra's Hybrid Theory of Induction (HTI), the core insights of Norton's critique can be fully accommodated within a Bayesian framework by fully recognising the role of models in Bayesian inductive logics.

The HTI provides a unified account of induction in which rules of inductive inference are articulated as sentence schemas whose accuracy depends on their ideal warranting conditions, that is, on specific factual conditions obtaining in the world. By reconstructing Bayesian inductive logics within this framework, I show that Bayesian rules are not freestanding formal principles, but rules whose correctness depends crucially on the accuracy of the underlying probability models. This dependence mirrors Norton's central claim that inductive support is ultimately grounded in material facts.

On this view, Bayesian inductive logics yield accurate information about confirmation only insofar as their probability models accurately represent the target system. As model accuracy degrades, so does the truthlikeness of the information about inductive support delivered by Bayesian confirmation measures. This result illuminates the specific epistemic role of model accuracy in Bayesian inference and provides a principled justification for practices such as model-checking.

The reconciliation offered here shows that one can remain a Bayesian while fully embracing the spirit of Norton's materialism about induction. In this sense, Norton and Bayesians can, after all, agree: Bayesian inductive logic and material theories of induction are not rivals, but complementary perspectives once embedded in the HTI framework.

Shelley Cameron – **Artificial Intelligence from Sociotechnical and Eco-Cognitive Perspectives**

This paper analyzes artificial intelligence (AI) through the frameworks of sociotechnical imaginaries (STI) and eco-cognitive theories of reasoning. Sociotechnical imaginaries, as developed in Science and Technology Studies, describe collectively held visions of desirable and undesirable technological futures that shape public understanding, political support, and ethical evaluation of scientific change. Literary and cinematic representations, including Frankenstein to science fiction robotics and popular film, play a significant role in forming these imaginaries, influencing how AI is imagined as either beneficial or threatening. The paper complements this perspective with Magnani's eco-cognitive framework, which understands cognition as emerging from interactions among humans, artifacts, and environments. Examples such as AlphaGo illustrate both the power and limitations of contemporary AI within constrained domains. While STI foregrounds the cultural and political dimensions of AI, the eco-cognitive approach emphasizes its cognitive and ethical implications. Together, these perspectives provide a more comprehensive account of AI as a sociotechnical and cognitive phenomenon.

Son Se Hoon – **Residuals as Style: Re-reading Manfred Frank through Abductive Discretization and the Residual Ledger**

Model-based reasoning requires discretization: it converts a continuous world into variables, states, thresholds, and categories. Yet discretization is constitutive, and therefore it necessarily produces residuals—near-miss alternatives, uncertainty spikes, counterfactual fragilities, and even "unsayable" aspects of cases that the schema cannot represent. Standard practice treats such residuals as noise to be eliminated. This paper proposes the opposite orientation, grounded in Manfred Frank's reconstruction of early German Romanticism and his critique of neo-structuralism.

Following Frank's account of infinite approximation, no model should be mistaken for closure: understanding proceeds through revisable stages rather than final possession. Following his critique of structural capture, subjectivity and meaning are not exhausted by formal codes; what is most decisive often appears where general form fails. Drawing on Frank's philosophical treatment of style, residuals are reinterpreted as structured signatures of singularity: not random defects, but "style-like" traces of non-subsumption produced by discretization itself.

On this basis the paper introduces Abductive Discretization (AD). While abduction in MBR is often framed as hypothesis generation for surprising facts, AD treats residuals as surprising facts about the schema: they expose conceptual poverty in the representational vocabulary. The abductive move is therefore not only to search for better hypotheses within a fixed code, but to revise the code so that residuals become intelligible—yielding an open-schema dynamics consistent with Romantic non-closure. Thus, AD extends MBR from reasoning within a model to reasoning about the model's own boundaries: residuals are treated as model-generated prompts for revising the scheme of representation rather than as mere noise.

To sustain this dynamics, the paper proposes the Residual Ledger (RL) as technical hermeneutics: a structured memory for residual traces that preserves what the model renders unsayable, enabling accumulation (patterned exclusion), contestation (re-description), and disciplined schema revision. Brief vignettes—safety that becomes captivity, representation that erases ordinary faces, and urban classification across Singapore, Seoul, and Busan—illustrate residuals as the motor of conceptual change. The central claim is that residuals should not be erased as failures; they are the very site where model-based reasoning remains open to singular life.

Srivastava Zachary – **Representing Through Analogical Inference**

In scientific modeling, there is an ongoing question regarding how models represent their targets. This paper draws on Mary Hesse's (1966) view of analogies as constituting the model-world relationship and combines it with Mauricio Suarez's (2004) inferential conception of representation. There are three steps in doing so. First, I define analogies in terms of similarity among the explanatory dependence relations in the source and target domain. Second, I consider Suarez's arguments against similarity accounts of representation and how the agent's mismatched familiarity with the source and target domain generates representation force. This provides the leverage for the third step, which demonstrates how restricting the similarity in analogies to the explanatory dependence relations and acknowledging the asymmetrical familiarity with the source and target domain culminates in a non-naturalistic theory of representation that relies on analogical inference.

Sterpetti Fabio – **Counterfactuals, Models, and Logical Expressivism**

In recent years, attempts have been made to develop inferentialist and expressivist accounts of scientific explanations. These challenge the traditional, representationalist view of scientific explanation, which has been criticised in recent decades for its inadequacies. One of the main difficulties the representational view faces is accounting for the role idealisations play in scientific explanation. Indeed, many authors have argued that highly idealised models, in which idealisations play an indispensable role, provide genuine scientific explanation and increase our understanding, even if they fail to accurately represent their targets. A debate that was related to this issue was about whether understanding can be reduced to knowledge and if it is factive. Those who proposed that understanding is not reducible to knowledge reinforced the perspective of developing a sophisticated anti-realist view in the philosophy of science. Most of the proposed inferentialist and expressivist accounts of scientific explanation are based on Brandom's view of logic. However, more sophisticated formulations of logical expressivism have recently emerged that could offer useful insights to those aiming to refine an expressivist position in the philosophy of science. For example, some of those formulations of expressivism seem promising in accounting for some aspects of counterfactual reasoning in science. Indeed, when it comes to scientific explanation, the issue of model-based reasoning cannot be avoided, and this type of reasoning is closely associated with counterfactual reasoning. When dealing with highly idealised models, one deals with counterfactuals that have a metaphysically impossible antecedent, i.e., counterpossibles. Some philosophers of logic have developed expressivist semantics for counterpossibles in logic, or counterlogicals, by building on Einheuser's attempt to provide a conventionalist account of counterfactual reasoning. It will be argued that this proposal can adequately account for certain aspects of the process of discovery in a way that is consistent with the notion of non-factive understanding.

Thompson Bruce – **Defining Reversal Negation Using Bipolar Fuzzy Sets**

Logicians have considered three understandings of logical negation: complementation negation, subtraction negation, and reversal negation. Only complementation negation—the type of negation used in truth-functional logic—has so far been given a rigorous definition using set theory. This paper presents an extension of fuzzy set theory under which fuzzy membership values are not single values but a range of values either greater than or less than an initial value. This concept of "polarized" membership values is then used to define reversal negation. The logical operators 'not' and 'if…then…' may then be defined in a way that gives Aristotle's Thesis, $\sim(\sim A \rightarrow A)$, the status of a logical tautology. Logic systems that include Aristotle's Thesis as a theorem—known as "connexive" logics—model the way common speakers use negation and entailment better than truth-functional systems do.

Trujillo Joaquin – **Dark Intelligence**

This article proposes supplanting the sobriquet – "artificial intelligence" (AI) – with "dark intelligence" to offset common presuppositions belying the technology commensurate with a

disregard of the originary (ursprünglichen) comprehension of being (Seinsverständnis). A hermeneutic-phenomenological perspective is enacted. The report discerns the comprehension of being – the pre-thematic understanding of "is" (vorontologischen Seinsverständnisses) – as the ownmost (Wesen) of human intelligence, identifies inconsistencies in the "artificial intelligence" appellation correlated to Dasein's factical (faktisch) obliviousness to the existential (existential) (constitutive moment of human being), and lays out the case to replace the "artificial intelligence" moniker to clarify the way AI is thought.

Urbanski Mariusz; Aleksandra Wasielewska – **Assessing Explanation Quality Without Ground Truth: A Multi-Layered Approach to Abductive Reasoning in Underdetermined Scenarios**

Contemporary accounts reconceptualise abduction as a compound reasoning process encompassing hypothesis generation, iterative testing through controlled information acquisition, and eventual stabilisation as a structured representation of events. In open-ended abductive tasks, however, multiple competing explanations may simultaneously be mutually incompatible, internally coherent, and evidentially defensible. This raises a fundamental methodological challenge: what principled criteria should distinguish better from worse explanations when no single correct answer exists? We address this challenge using the Find Out paradigm, which elicits full narrative explanations in response to ambiguous investigative scenarios whilst recording both the final explanatory products and the epistemic processes by which they are constructed. Our dataset comprises 134 written reports generated for a deliberately underdetermined scenario involving a police stop of a public figure, in which critical elements are deliberately left unspecified. We propose a three-layer evaluative framework. The first layer assesses structural adequacy: whether reports explicitly address the core narrative gaps in the scenario. The second layer targets coherence through calibrated human judgment, operationalising dimensions such as internal consistency, narrative integration, and disciplined epistemic reasoning. The third layer encodes each report as a situational graph (that is, a directed acyclic graph representing situations and temporal-sequence relations), enabling quantitative analysis of formal structural properties. Data collection is complete, and all reports have been encoded as situational graphs. The present talk focuses on the first and second layers, for which research is ongoing, presenting preliminary findings from structural gap-filling analysis and the iterative development of coherence dimensions. This framework offers a principled methodology for evaluating abductive explanation quality in naturalistic settings, bridging qualitative and quantitative approaches to model-based reasoning and advancing understanding of how humans construct and communicate explanations under genuine epistemic uncertainty.

Viola Marco; Valentina Petrolini – **Affective Sailors: The Affect Circumplex as a Model for Emotional Regulation**

Affectivity is widely regarded as a core component of mental life, underlying emotional episodes, moods, and feelings. One influential framework for characterizing affect is the Affect Circumplex, which maps affective experience along two bipolar dimensions – i.e., arousal and valence. Although this model is extensively used in psychological research and measurement, its theoretical commitments are often left implicit. In this talk, we distinguish between a realist interpretation of the Circumplex – one that is committed to the ontological reality of arousal and valence as components of "Core Affect" – and an instrumentalist interpretation, which treats these dimensions as theoretically useful but non-ontologically loaded constructs. We first offer reasons for skepticism toward the realist reading, drawing on concerns about the neural basis of Core Affect, the existence of mixed affective states, diachronic affective dynamics, and the complex relationship between arousal and valence. Against a deflationary rejection of the model, we build on our previous work to show that arousal and valence interact over time. Arousal modulates the intensity of felt valence through a process we call magnification, accounting for the tendency of high-arousal states to co-occur with extreme positive or negative valence. We then extend our account by interpreting arousal as a proxy for salience allocation rather than a causally autonomous entity. This interpretation is in line with

appraisal theories of emotion, but also makes room for an instrumentalist reading of the Affect Circumplex. On our view, the Circumplex does not merely work as a snapshot of current affect, but as a dynamic framework for affect regulation. To better capture this idea, we introduce a guiding metaphor: affective agents are sailors navigating the troubled waters of their affective lives. While sailors cannot control the wind (i.e., external events), they can adjust their sails – modulating salience, engagement, and arousal – to steer toward more desirable affective outcomes. Through this metaphor, we analyze a novel regulatory mechanism we call sheltering, understood as the conscious or unconscious lowering of arousal through salience disengagement. We show how sheltering can be adaptive or maladaptive, with implications for both everyday affect regulation and clinical conditions.

Visokolskis Sandra; Graciela Marta Chichi – **How Does Peircean Abduction Model? A Controversial Historical Argument Presumably Appealing to Aristotle**

The paper aims to propose an abductive modelling approach in the context of a search for scientific explanation in creative discoveries, based on a variation of Peirce's multifaceted notion of abduction called 'transduction', that adds two preliminary stages to the reasoning from effect to cause, what we would properly call abductive explanation: (1) a resemblance stage, where similar cases are evoked to proceed by likeness, and (2) an analogical transfer, to complete the structuring of the innovative process. The use of resemblances before giving a properly abductive explanation brings with it the advantage of emphasizing the degree of plausibility that would make abduction capable of producing new information, even despite the elusive, non-deductive manner in which it is carried out. Given that the purpose of Peirce in the first place was not to validate abduction, but rather to show his cognitive versatility when introducing novelty, this work is interested in discussing the possible historical bases that Peirce would have rescued from Aristotelian texts where the theme of similarity is central. In this sense, in 2014, Jorge Alejandro Flórez Restrepo published an emblematic article in the Transactions of the Charles S. Peirce Society journal, where he raises two points: (i) Peirce's theory of the origin of abduction in APr. II.25 is mistaken, since Aristotle does not discuss deductively invalid second-figure syllogisms there, which is the case with abduction. And (ii) Peirce should have relied on APo. I.13, in his search for the Aristotelian origin of his notion of abduction. However, in a 2019 article in the same journal, Francesco Bellucci accepts point #1 but rejects point #2. As a counter-response to both Flórez (2014) and Bellucci (2019), this paper seeks to solve these issues from a diametrically opposed point of view. The intention here is not to disagree with Flórez and/or Bellucci, but rather to address more cognitive than logical issues, which are better solved through a model of abduction based on resemblances and analogies.

Zennaro Fabrizio – **GDP and Economic Performativity: From Counterperformativity to Theoretical Responsibility**

This study examines the role of Gross Domestic Product (GDP) within contemporary economic debate through the lens of economic performativity. It argues that GDP is not a neutral metric but actively shapes and reinforces normative assumptions embedded in the neoclassical theory tradition. By highlighting these performative effects, the research calls for explicit recognition of the broader implications of economic theories and advocates for greater awareness and accountability in their application.

Zhang Zhitao – **A Significance Test for Causal Inference**

The literature on causal inference focuses on recovering causal structure from the population probability, while neglecting that the stochastic nature of data may lead to an unfaithful presentation of population probabilities. Statisticians developed significance tests to evaluate this problem. In this paper I explore whether the basic ideas underlying such tests may also be applied to similar problems in the context of causal inference. I point out problems for Fisher's approach, propose a prototype significance test inspired by Rubin's work, and discuss whether this prototype test may solve underdetermination problems from the causal search literature.

Symposia

Models of Serendipity. How to Map the Role of Chance in Discovery

Arfini Selene; Matteo Costa

This workshop aims to explore and propose models for understanding the role of serendipity in scientific discovery, examining the interplay between chance, the "prepared mind" of scientists, and the cognitive and epistemological processes that make such discoveries possible. Traditionally defined as the finding of the unexpected through both luck and sagacity, serendipity encompasses a cluster of processes that remain under-theorized despite their centrality in the history of science. By bringing together perspectives from the philosophy and history of science and technology, the workshop aims to advance a systematic inquiry into how this phenomenon can be modeled, while opening a dialogue on its broader meanings and implications. Theories and ideas presented in this symposium will focus on models that account for both the prospective openness of serendipity, which makes scientists anticipate multiple possible outcomes and create space for surprises, and retrospective reinterpretation, where past data can be reframed in light of new insights. Scholars will also discuss the role of counterfactual reasoning, which helps scientists reconstruct alternative pathways of discovery and better understand the role of contingency in scientific practice. In this sense, models of serendipity will clarify the mechanisms through which the unexpected is not merely noticed but transformed into a resource for knowledge growth and production. The symposium will also draw attention to the technological and material dimensions of discovery, calling into play the concept of affordance, which can be defined as a bodily and mental action possibility provided by tools, environments, and artifacts. Indeed, experimental apparatuses, digital platforms, and designed environments all provide opportunities for productive surprises. The philosophy of technology, therefore, will also be essential for understanding how serendipity is materially and structurally mediated, as well as how conditions can be created to favor its occurrence.

Copeland Samantha – **Cave Exemplar – Modeller [of Serendipity] Beware!**
As an emergent phenomenon, serendipity is difficult to model. I will address in this talk a number of difficulties and a proposal for modelling despite them. First, the tripartite definition of serendipity normally offered includes three distinct kinds of factors: the cognitive factor of noticing or recognizing potential value and sagacity, the supporting environmental factors that enable the process of discovery, and the 'value' of the discovery that enables us to call this a 'happy' event that was more than just luck. A further factors is that the emergence is both epistemic and ontological, give the chance factor in serendipity. First, give the arbitrary nature of identifying necessary factors in a discovery process (see, eg. Fleck), it is difficult to say if even retrospective models capture the right details. Second, only indicators and indirect causal factors can be captured in prospective or predictive models, thus preventing certainty about their identification and usefulness in new contexts. Nonetheless, I argue that modelling is a worthy exercise and that anecdotes and situational descriptions can be enough to assess whether a change will increase potential serendipity in an environment, or what skills people can cultivate to increase serendipity in their lives. I suggest that contexts such as ethical situations, transdisciplinary settings, and observational settings (eg. Astronomy, as well as model-building itself) are naturally serendipity-rich, and they can be modelled to show this. Indicators, in turn, can be formulated via analogy—situations such as the above are those that encourage psychological safety, pluralist ontologies and effectual reasoning, for example. Exploring serendipity through this exercise, finally, provides grounding for a generalized understanding of the kind of reasoning that enables tripartite processes such as serendipity.

Costa Matteo; Stefania Iaria; Federica Giordano; Selene Arfini – **When Anomalies Become Possibilities: Serendipity and the Emergence of New Affordances in Scientific Discovery**
The philosophy and history of science have so far paid insufficient attention to an issue that concerns both the dynamics of discovery and the role of tools in scientific research: the occurrence of serendipitous discoveries and the emergence of new affordances. Notwithstanding the significant characteristics they share, indeed, the concepts of serendipity

and affordance have consistently been analyzed and studied separately, as distinct and unrelated ideas in their respective literatures. Yet, this paper argues that within their relationship lies the key to understanding how serendipity can bring about a dramatic change in perspective for the involved researchers and provide the opportunity to discover and explain new phenomena. In particular, we maintain that, when a serendipitous discovery process arises, it implies the actualization of a new affordance that co-emerges and co-evolves with the new knowledge deriving from the discovery itself. The new affordance and the Gestalt-switch, then, emerge simultaneously as interdependent aspects of the same transformative process, both characterized by relational and dynamic dimensions.

Furlan Stefano – **Seeker of the Larger View: John Wheeler and His Heuristics, Between Nuclear Fission and Black Holes**
John A. Wheeler (1911-2008) stands as a towering figure in 20th-century physics, contributing to numerous fields and stimulating countless results through his mentorship. He also left an extraordinarily rich archive (now housed at the American Philosophical Society Library in Philadelphia), which includes correspondence, drafts, excerpts, and notebooks in which he recorded his ideas, conversations, interests, and conference notes over several decades. In a highly rigorous discipline where attention typically focuses only on crystallized, published results, these unique sources allow us to trace in remarkable detail the gradual maturation of concepts and the metamorphosis of research questions. Equally fascinating is that, in the mature phase of his career, Wheeler himself began explicitly meditating – with notable originality – on the conditions of scientific creativity, how to foster it, and on "missed opportunities". Recently, Furlan (2022) attempted to characterize a more flexible and enriched notion of "pursuitworthiness" by examining the crucial decade of Wheeler's black holes research. The aim was to shift the focus toward the heuristic dimension – broadly conceived, with its indissoluble intertwining of internal and extrinsic factors – to develop a model accounting for how scientific discoveries emerge from the dynamic interplay between contingency and intellectual readiness. This approach demonstrates that, in well-documented cases such as Wheeler's, the creative process is rigorously investigable and does not reduce to some unfathomable mystery of individual genius. I will thus compare this proposal – supported by documentation from Wheeler's notebooks – with recent literature on an ecological approach to serendipitous discoveries (Arfini & Costa, 2025; Copeland, Ross, & Sand, 2023): by examining how Wheeler conceptualized both prospective openness and retrospective reinterpretation, I will explore the convergences and tensions between these frameworks. Finally, I will introduce Wheeler himself as an interlocutor, drawing on his reflections on the quasi-discovery of nuclear fission to illuminate his own understanding of scientific creativity and missed opportunities. Indeed, this historical case study offers methodological insights into how archival evidence can ground philosophical accounts of discovery processes.

McClelland Thomas – **Salience, Skill and Asking The Right Questions**
Questions play a central role in structuring scientific inquiry. But how are scientists able to arrive at the right questions? This paper explores the role that salience plays in question-asking practices in science. The questions we ask are often prompted by what pops out to us i.e. what we find salient. A scientist can thus arrive at the 'right' questions if they find the 'right' things salient. But what enables the right things to become salient? This paper considers two complementary answers. 1) Salience and Skill - The empirical literature strongly suggests that skilled action is underwritten by salience structures that attune an agent to objects and features relevant to their task. This is supported by findings in domains such as playing chess, playing football or assessing a medical biopsy. However, in these domains it is relatively clear what makes something relevant to the agent's task but in scientific inquiry relevance is much more nebulous. I argue that scientific inquiry this is more analogous to our engagement with art and suggest that empirical findings regarding salience in skilled art engagement can shed light on scientific enquiry. 2) Salience and Models - One of the virtues of a good model is that it can make the right things salient. For example, data that previously seemed innocuous can suddenly become salient in the light of a particular model. This then prompts the inquirer to explore why that data obtains. A model can thus be regarded as a tool for making things salient. Accordingly, a good model is one that makes the right things salient and thus prompts the right questions. Crucially, this virtue does not depend on models bearing a representational relationship to the world.

Convergence Through Modeling

Nappo Francesco

Is the fact that a given prediction holds across distinct scientific models evidence for that prediction? In the philosophical literature, this question has been discussed under a variety of labels, including robustness, consilience, convergence, triangulation, and stability. In what follows, we refer to the relevant phenomenon as convergence through modeling. Determining whether, and under what conditions, convergence through modeling has evidential force lies at the core of many debates in contemporary philosophy of science.

Such debates arise across a wide range of scientific domains. In the philosophy of climate science, for example, there is ongoing disagreement about whether convergence among long-term simulations produced by distinct climate models can serve as evidence for climate predictions. In the philosophy of biology, related concerns arise in discussions of whether agreement among models of organisms that rely on different simplifying assumptions can support claims about underlying causal mechanisms. Similar issues also emerge in physics and the social sciences, where model-based reasoning plays a central epistemic role.

The expectation that different instances of convergence through modeling share a common argumentative structure has motivated the search for general, top-down epistemological principles governing scientific modeling. Prominent proposals in the philosophical literature appeal to conditions such as model independence or the stability of predictions across sufficiently similar models. These approaches rest on the assumption that there exist general epistemological principles that regulate the evidential use of models across scientific fields.

By examining case studies of convergence through modeling across different disciplines and research contexts, this symposium addresses two main questions: whether there are sufficient similarities across cases of convergence to warrant a unified epistemological account, and whether principles such as independence and stability are plausible as regulative norms for scientific modeling. For the variety of case-studies and viewpoints considered, this symposium aims to serve as a reference point for future philosophical reflection on the epistemological role of convergence in model-based reasoning.

For the variety of case-studies and viewpoints considered, this symposium aims to serve as a reference point for future philosophical reflection on the topic of convergence in the scientific domain.

References

Castellani, E. 2024. Convergence strategies for theory assessment. Stud Hist Phi Sci, 104:78-87.

Fletcher, S. 2020. The principle of stability. Phil Imprint. 20:3

Frigg, R. et al. 2013. The myopia of imperfect climate models. Phil Sci. 80(5):886-897.

Knuuttila T. & Loettgers, A. 2011. Causal isolation robustness analysis. Biol & Phil. 26(5):773-791.

Levins, R. 1966. The strategy of model building in population biology. Amer Sci. 54(4):421-431.

Odenbaugh, 2011. True lies: realism, robustness, and models. Phil Sci. 78(5): 1177-1188.

Parker, W. 2011. When climate models agree. Phil Sci. 78(4): 579-600.

Snyder, L. 2006. Consilience, confirmation, realism. In Achinstein, P. (ed.) Scientific Evidence.

Wimsatt, B. 1981. Robustness, reliability, overdetermination. In Brewer, M. (ed.) Scientific Inquiry.

Winsberg, E. & Goodwin R. 2016. The adventures of climate science. Stud His Phil Sci. 54: 9-17.

Cangiotti Nicolò – **What Happens Near the Edge of a Black Hole?**

Next, mathematical physicists Nicolò Cangiotti will present on the philosophical aspects of convergence in the emerging field of black hole thermodynamics. His talk, titled 'What happens near the edge of a black hole?', will examine the appearance of so-called Hawking radiation across gravitational, analogue, and dispersive models, arguing that this convergence is best understood as structural robustness rather than as evidence for universality, as defended in recent philosophical literature. What unifies these models is the presence of horizon-like causal features that enforce a mismatch between 'in' and 'out' representations, rather than a shared dynamics. His case is important to the general question of the symposium

because it shows how convergence across models may support analogical inference while falling short of warranting claims of universality.

Castellani Elena; Emilia Margoni – **Cluster Transfer and Non-Independence**
Next, Elena Castellani and Emilia Margoni will discuss whether independence between models is required for the evidential value of convergence with their talk 'Cluster transfer and non-independence'. By referring to the concept of cluster transfer, a generalization of the unit of knowledge transfer across models, they will argue that a common result can follow from this form of transfer in a weakly robust way, that is, without necessity of independence. They will illustrate this claim through two cases of cross-fertilization between quantum field theory and condensed matter physics: the birth of renormalization group theory and the introduction of the idea of spontaneous symmetry breaking in particle physics. In both cases, the transfer of a cluster of ideas, formal tools and methods played a crucial role for the acceptance of these theories and their further development.

Giordani Alessandro; Francesco Nappo – **Stability Reconsidered**
Next, 'Stability reconsidered' by Alessandro Giordani and Francesco Nappo will problematize Fletcher's (2020) principle of stability, according to which an inference from a model is justified only if it remains stable under small variations of modeling assumptions (or in 'sufficiently similar' models). They will scrutinize the formal and conceptual limits of the proposed notion of 'similarity', discussing the idea that similarity relations are underdetermined and purpose-relative. They will further note that convergence across non-stable but diverse models is sometimes sufficient for justification via routes that do not depend on local stability. Finally, they will question the scope of the proposal: many modeling practices pursue epistemic aims, such as the generation of scenarios in climate change research, to which the principle of stability does not apply. One upshot of their talk will be that the justificatory import of models exceeds the cases covered by the principle as proposed by Fletcher.

Knuuttila Tarja; Andrea Loettgers – **Dependent by Design, Independent by Realization: Multiple Models of Circadian Clock**
Tarja Knuuttila and Andrea Loettgers will kick off the symposium with their talk 'Dependent by design, independent by realization: multiple models of circadian clock'. They will distinguish convergence in the 'independent determination' sense and in the 'causal isolation' sense and examine how these strategies are combined in modeling the circadian clock. In this field, mathematical models are integrated with synthetic and electronic models, as well as model organisms. While these models are not independent by design, they exploit independent media – mathematical, digital, and biological. Through this case, the authors will highlight the limits of top-down accounts that treat convergence across models as a general source of evidential support, showing instead how its role is sensitive to aims, material realizations, and systems under study.

Abduction and Model-Based Reasoning in Action

Park Il Memming; Gonzalo G. de Polavieja; Ábel Ságodi; Woosuk Park; Doohyun Sung

We are in the heyday of abduction and model-based reasoning in logic, philosophy of science, artificial intelligence, cognitive science, and many other scientific disciplines. But to what extent are practicing scientists influenced by this trend? How would they describe what happens in their labs in abductive terms, and where—if anywhere—could explicit attention to abduction and model-based reasoning accelerate scientific discovery? How should this reshape both philosophical analysis and the design of AI systems for science?

We discuss such interdisciplinary questions by sharing our reports from two neuroscience laboratories. We draw from Lorenzo Magnani's (2009, 2022) works on abduction, Nancy J. Nersessian's (2022) works on interdisciplinarity, and the philosophical literature on interventionism (Hacking, 1983; Westerblad & De Regt, 2025), haptic realism (Chirimuuta, 2021, 2023, 2024), and analogical reasoning (Hesse, 1963) to construct our analytical framework.

Within the framework, we explore three topics. First, we begin by examining abduction as it is enacted in contemporary laboratory practice, drawing on comparative reports from the two laboratories. These cases show that abduction rarely appears as a single inferential step.

Instead, it unfolds through the construction, constraint, and revision of model spaces over time, tightly coupled to experimental intervention. In one case, abductive reasoning is embedded in Bayesian and statistical modeling loops, where hypotheses are graded rather than eliminated and where closed-loop experimental design actively shapes what counts as evidence. In the other, abductive progress emerges from iterative interaction between simplified theoretical models and targeted experiments, including the use of non-statistical modeling formalisms aimed at capturing algebraic structure. Across both cases, abductive success depends on the ongoing reconfiguration of representational and experimentally imposed constraints.

Secondly, we broaden the analysis to creative and manipulative abduction in biological and artificial agents. In biological systems, learning often exhibits sparse but decisive transitions, as if a new internal model or rule has been acquired. Such transitions appear in non-human primates during cognitive learning and in collective animal behavior, where individuals seem to infer unobserved causes from indirect cues and the behavior of others. Artificial agents sometimes show analogous performance jumps, but these typically arise from gradual parameter updates rather than explicit hypothesis formation. Comparing biological and artificial cases helps clarify which aspects of creative and manipulative abduction depend on embodied exploration, intervention, and environmental engagement—and which can be approximated by current and future AI agents capable of creative and manipulative abduction for scientific discovery.

Finally, Doohyun Sung and Ábel Ságodi analyze ideal modes of collaboration between human researchers and an "AI co-scientist," with particular attention to the distribution of cognitive labor in model-based reasoning, the design of AI models, and the contribution of analogical reasoning. By committing to interventionism and haptic realism, Doohyun and Ábel propose that science is best understood as an interventionist practice rather than data-to-theory inference simpliciter.

In this context, scientific models render complex systems intelligible by suppressing inessential details and foregrounding (causal) structures, thereby supporting counterfactual reasoning, prediction, and control. What follows is that an AI scientist should not aim to merely maximize predictive accuracy, but to construct simplified, ontology-bearing models that can be experimentally probed and revised. Regarding well-known issues in AI including epistemic opacity and distortion of phenomena due to abstraction, some have argued that AI systems used in scientific practice must be adequately designed to support causal and mechanistic understanding, not just statistical prediction; others have argued that abstractions should be evaluated relative to their epistemic aims, i.e., prediction, explanation, or manipulation, rather than parsimony alone. Doohyun and Ábel argue that analogical reasoning plays a crucial supporting role in optimizing the overall collaborative process. For one thing, theories require models and material analogies to connect formal structures to the physical world, lest they devolve into "formal games" without empirical grip. For another, cognitive science has shown that human understanding relies on structured representations grounded in informal ontologies of objects, processes, and mechanisms, which AI systems may not possess by design. Therefore, the AI scientist, via disciplined analogical reasoning, could scaffold abstraction, thus anchoring computational models to human-understandable domains while preserving their capacity for prediction, manipulation, and control.

In sum, our symposium establishes a foundation for interdisciplinary discussion of abduction and model-based reasoning in action, clarifying how philosophical analysis and AI–neuroscience research can mutually inform one another at the levels of ontology, epistemology, and methodology.

References

Chirimuuta, M., 2021. "Prediction versus understanding in computationally enhanced neuroscience." Synthese, 199(1), pp.767-790.

Chirimuuta, M., 2023. "Haptic realism for neuroscience." Synthese, 202(3), p.63.

Chirimuuta, M., 2024. The brain abstracted: Simplification in the history and philosophy of neuroscience. MIT Press.

Hacking, I., 1983. Representing and intervening: Introductory topics in the philosophy of natural science. Cambridge University Press.

Hesse, Mary B. 1963. Models and Analogies in Science. University of Notre Dame Press.

Magnani, L. 2009. Abductive Cognition: The Epistemological and Eco-cognitive Dimensions of Hypothetical Reasoning. Springer.

Magnani, L. 2022. Discoverability: The Urgent Need of an Ecology of Human Creativity. Springer Cham.
Nersessian, N.J. 2022. Interdisciplinarity in the Making: Models and Methods in Frontier Science. MIT Press.
Westerblad, O. and De Regt, H.W., 2025. "Scientific Understanding Beyond Representing: Lessons from Ian Hacking's Work." The Monist, 108(4), pp.353-371.

de Polavieja Gonzalo G.; Woosuk Park; Il Memming Park – **Creative and Manipulative Abduction in Animals and AI Agents**

Park Il Memming; Gonzalo G. de Polavieja; Woosuk Park – **Abduction in the Wild: A Case Study in Neuroscience and Biology Labs**

Sung Doohyun; Ábel Ságodi – **Prospects of AI Scientists in Neuroscience: Analogical Reasoning in Focus**

Machine Learning Meets Molecules. Philosophical Reflections on AI-Driven Chemistry and Biochemistry

Seifert Vanessa

In recent years, expanding applications of machine learning (ML) and artificial intelligence (AI) have seen changes to both the way that science is practiced and even, potentially, its epistemic foundations. Going beyond the ethical challenges raised by these machines, there are a number of philosophically interesting consequences of applying AI to scientific practice. For instance, we can ask whether tools from AI/ML have the capacity to improve our understanding of a target system, or whether such tools place pressure on traditional accounts of scientific realism. To answer these general questions, we think it fruitful to investigate the application of specific AI methods within specific contexts of inquiry.

Among other domains, AI models are used in scientific practice to further study molecular cases. This involves approaching molecules at different scales, from the chemistry of elements to chemical substances and macro-molecules such as proteins. These programs are projected to have high scientific value. For instance, the program AlphaFold 2, which predicts proteins' structure from sequences, is said to be one of the most revolutionary applications of artificial neural networks. The relevance of these cases to philosophy has by now been recognised, with AI offering a bountiful source of input for philosophical analysis. Recent scholarship on the philosophy of AI/ML in science has focused on some of the following questions: Does the opacity or value-ladeness of ML models undermine their efficacy or their objectivity? Can ML models both predict and explain phenomena? Which kind of representations are present in ML models? What are ML models discovering, and is this a contribution to core scientific questions? Do the discoveries made with the aid of AI tools challenge how we think about scientific realism?

In the symposium "Machine learning meets molecules: Philosophical reflections on AI-driven chemistry and biochemistry" , we will explore these questions in light of applications of AI to molecular science. We believe that research of this kind stands to be philosophically productive for the following reasons: i) applications of AI to molecular prediction and discovery has been considered one of the core scientific developments of recent years; ii) this research provides us with a concrete and case-sensitive analysis of the philosophical implications of AI; iii) molecules are a core bridge between the physical and biological realm, and a better understanding of them can provide new insights into our understanding of life.

To deal with these questions, our symposium will feature three papers, each of them touching upon core issues raised by AI at the molecular scale. Concluding, this symposium offers novel insight into the role that AI plays in molecular sciences and how these applications can contribute to core debates in philosophy of science.

References

Cabrera, F. (2021). The Fate of Explanatory Reasoning in the Age of Big Data. Philosophy and Technology. 34, 645–665 https://doi.org/10.1007/s13347-020-00420-9
Gerrard, W., Bratholm, L. A., Packer, M. J., Mulholland, A. J., Glowacki, D. R., & Butts, C. P. (2020). IMPRESSION–prediction of NMR parameters for 3-dimensional chemical

structures using machine learning with near quantum chemical accuracy. Chemical science, 11(2), 508-515.
Krishnan, A., Valeri, J. A., Jin, W., Donghia, N. M., Sieben, L., Luttens, A., ... & Collins, J. J. (2025). A generative deep learning approach to de novo antibiotic design. Cell, 188(21), 5962-5979.
Northcott, R. (2020). Big data and prediction: Four case studies. Studies in History and Philosophy of Science Part A 81 (C):96-104.
Ratti, E. (2020). What kind of novelties can machine learning possibly generate? The case of genomics. Studies in History and Philosophy of Science Part A 83 (C):86-96.
Rowbottom, D. P., Peden, W., & Curtis-Trudel, A. (2024). Does the no miracles argument apply to AI?. Synthese, 203(5), 173.

Bellazzi Francesca – **Does AlphaFold Solve the Protein Folding Problem?**
Dr Francesca Bellazzi, Durham University will then move to consider a case from biochemistry: AlphaFold 2, the algorithm able to predict proteins' structure from their sequences. This talk will investigate whether AlphaFold solves the problem of protein folding and whether, following its results, we are justified in supporting the reduction of protein structure to sequences. It will argue that this form of reductionism is not warranted and that protein structure is predicted thanks to the interactions of a complex set of data, comprising evolutionary and structural information.

Mallory Fintan; Mel Andrews **– Constructing Chemical Manifolds**
Dr Fintan Mallory, Durham University, and Dr Mel Andrews, Princeton University, will present a joint work, interrogating what is known as the manifold hypothesis in ML as it is born out in the chemical sciences. Their talk will explore several interpretations of the manifold hypothesis of varying ontological strength in light of a critical analysis of the apparent toroidal structure of the ethanol manifold. Should this mathematical structure be understood as real and wordly, a mere artefact of data handling, or a conceptual reification? This analysis raises questions, in turn, about the nature of scientific realism: in particular, does scientific realism entail a commitment to the truth-aptitude of our conceptual tools? Or merely the propositional content resultant from their instrumental usage?

Seifert Vanessa **– Could AI Undermine Standard Scientific Realism?**
Dr Vanessa Seifert, University of Athens will then consider whether AI challenges standard scientific realism. Atoms and molecules have constituted a paradigmatic case in defending the reality of unobservable entities. However, the way AI models are developed to predict novel chemical substances undermines the idea that the predictive success of science empirically supports the truth of its theories. This is because there are at least some AI models whose predictive success is not based on incorporating or following an underlying theory. To spell this out, Seifert presents two AI models that are developed to predict chemical substances: the generative intelligence framework developed by Krishnan et al. (2025) at MIT and the ML algorithm IMPRESSION developed by Gerrard et al. (2020) at the University of Bristol. These cases reveal two ways by which AI models are trained to make predictions, which in turn illustrate that predictive success in science is not always accounted for by the truth of an underlying theory. This is particularly problematic for the defence of scientific realism in terms of the No Miracles Argument which takes novel predictions in science to be best explained by the truth of our underlying theories.