GWP.2025

24-26 Mar 2025
Erlangen
Germany

Table of contents

Science, Skepticism, and Sociology: reframing the role of SSK in today's knowledge landscape

Morgan Adou * [1]

[1] Centre Gilles-Gaston Granger – Aix Marseille Université, Centre National de la Recherche Scientifique – France

This paper aims to critically evaluate the contribution of the Sociology of Scientific Knowledge (SSK) to discussions on skepticism, focusing on whether SSK can help counter science skepticism. At first glance, SSK has been accused of fueling skepticism itself. Classical critiques argue that by emphasizing the social construction of scientific facts, SSK promotes relativism, thereby undermining the objectivity and reliability of scientific knowledge. The first part of the paper will explore these critiques, examining how SSK has been seen as contributing to skepticism rather than counteracting it, and how Bloor (1976) predicted this ambivalence.
In response, I argue this critique misrepresents SSK's true capacity. Rather than promoting relativism, SSK offers a nuanced understanding of how scientific knowledge gains credibility through social processes. Instead of seeing the socially constructed nature of knowledge as a threat, SSK shows that the stability of scientific knowledge depends on institutional norms, collective practices, and shared rational criteria within the scientific community. By making these processes transparent, SSK demystifies science and can help rebuild public trust in scientific institutions, especially in the face of growing skepticism and populist critiques.

The third section will highlight how SSK's view of the revisability of scientific knowledge provides a constructive way to respond to science skepticism. One core claim of skeptics is that the inherent uncertainty of science undermines its reliability-how can scientific knowledge be trusted if it is always subject to change? SSK, however, frames this constant revision as a strength of science, not a flaw. By showing how scientific knowledge evolves through an iterative process involving empirical evidence and social consensus, SSK reveals how science progresses without claiming infallibility. This understanding can help reduce public fears about the provisional nature of scientific knowledge and explain how populisms are grounded in a teleological view of science (Bloor, 1976).

As a final argument, I will suggest that SSK is particularly effective against populism because it dispels this teleological view. Populist skepticism often arises from a misunderstanding that science moves toward an ultimate, immutable truth. By rejecting this vision, SSK frames science as a dynamic process, countering populist narratives that position it as elitist or authoritarian. Moreover, SSK enables a symmetrical explanation of populist skepticism without merely categorizing agents as irrational. By examining the social factors shaping both scientific knowledge and populist critiques, SSK offers a deeper understanding of why populism and skepticism arise,

*Speaker

rooted in the same social mechanisms governing all knowledge production.

In conclusion, this paper argues that while SSK has been criticized for promoting skepticism, it can serve as a powerful tool for addressing it. By rejecting teleological visions of science and explaining populist movements symmetrically, SSK offers a framework that counters populist skepticism without dismissing populist agents as irrational, making it a valuable resource for restoring trust in science today.

Keywords: ociology of Scientific Knowledge (SSK), Science Skepticism, Populism, Relativism, philosophy of social sciences

Patterns in Machine Learning and Science

Markus Ahlers * [1]

[1] Eberhard Karls Universität Tübingen = University of Tübingen – Germany

Machine learning (ML) is said to be able to recognize patterns in data sets that are not apparent to humans. In the area of supervised learning, this is shown by the fact that, for example, neural networks can recognize regularities, structures or patterns, on the basis of labeled data, on which reliable decision-making systems can be built. In the field of clustering, even the free recognition of patterns in data sets is highlighted as an achievement. The systems are even said to be able to detect patterns that are not recognizable to humans. This is one of the reasons why ML systems are also used in science.

In my talk, I want to link the handling of patterns in ML with the philosophy of science and thus show the extent to which human influence on pattern detection is also expressed by machine learning in the scientific process. In particular, this requires a clarification of the concept of patterns. Because even if the patterns found by ML systems are often used to optimize decision-making or classification in science, the exact epistemic status of the patterns in ML models remains unclear due to the lack of transparency of ML systems.

This is because it is unclear to what extent pattern recognition is also dependent on the assumptions of the developers or scientists. For example, even in unsupervised-learning, data preparation must be done. Before a method like KNN can be used, the data must be prepared in a vector space model. For this, it must be decided on the basis of which metrics the data points are placed in the vector space and how comparability of data points can be conceptualized. The better the metrics are selected, the better the features are determined, in other words, the better the data is prepared, the more certain it is that the ML-system will find the right patterns. So here also a human influence is needed to recognize patterns in the real world, which raises the question whether the values of the scientists influence the pattern recognition.

In my talk I focus on the question of what patterns actually are in scientific used ML-Models. I want to investigate the extent to which patterns in ML depend on human influence. In principle, the goal of the scientific use of ML is to find patterns that exist independently of researchers influence. In practice with its high demands on data preparation, the patterns seem to be dependent on assumptions by the researchers. Now there is already a discussion in the philosophy of science about realistic positions of patterns (Denett 1991, McAllister 2009, McAllister 2010), but so far this has hardly been connected with the methods of ML. Likewise, no connection was drawn to the previous epistemological investigation regarding methods in computer science, which are characterized by an explorative nature (Schiaffonati 2016). I would like to explore the possible points of connection and, if possible, use them to sharpen the epistemic understanding of ML.

Keywords: Machine Learning, AI, Patterns, Values

*Speaker

Transforming Standards: Observation contextualised

Sarwar Ahmed *† [1]

[1] University of Wuppertal – Germany

Despite its central epistemic role, the nature of the standards by which an observational claim is accepted has received little philosophical and historical attention. Chalmers (2013) has noted the transition from an Aristotelian doctrine of observation based on perception to the acceptance of telescopic observations promoted by Galileo's demonstrations. Dawid (2021) has noted the need to integrate meta-empirical assessments into the concept of empirical confirmation in the period of the acceptance of atomism that led to a justified discovery of microphysical objects. However, these authors have focused on specific historical periods without intending to provide a broader understanding of the implications of these transformations.
In this paper, I defend two claims: first, that an atemporal set of standards that can be used as necessary and sufficient conditions for accepting observational claims is alien to the nature of scientific practice. Second, without the integration of meta-empirical assessments developed by Dawid (2013, 2018, 2021) into the scientific method, the extension of the concept of observation to non-perceptual instrumental observation and microphysical entities would not have been possible due to the problem of unconceived alternatives for the source of the collected data.

When introducing a new observational method, one must address two epistemic problems: justifying the reliability of the data collected (calibrating and testing the instrument) and justifying the causal correspondence between the source and the data collected (addressing the problem of underdetermination of the source by the data). The philosophical literature on the acceptance of new methods of observation such as telescopes and microscopes is for the most part divorced from the influence of the acceptance of new theories. I argue that, historically, the acceptance of a new observational method is deeply intertwined with the acceptance of a new theory, resulting in the emergence of new standards of observation.

This paper examines the historical and epistemological components of observational standards in two major historical episodes in science, Copernicanism and atomism. It will be argued that assuming the reliability of Galileo's telescope, Thomson's microscope, Perrin's ultra-microscope and even his multi-methodical strategy were not sufficient for the community to immediately accept the observations. That is, the mere saving of the phenomena is not sufficient to establish conclusive claims about the existence of the source. A further step, namely a reliable inference to the source (treatment of the underdetermination of the source) was needed. I argue that in these two cases, Copernicanism and atomism, the standards of observation were transformed. In the former, observation was extended from naked-eye observation to non-perceptual instrumental observation. In the latter, the observation was extended to the observation of microphysical

*Speaker

†Corresponding author: sarwarmuhamad84@gmail.com

entities. Moreover, I argue that, in both cases, the meta-empirical assessments developed by Dawid (2013, 2018, 2021) played an essential role in the treatment of the underdetermination of the source.
This indicates that, in general, the rational mechanism that governs these transformations is a mutual relation between the reliability of observational method and the trustworthiness of the theory that predicts the phenomenon to be observed in any given context.

Keywords: Observation, Underdetermination, Transformation, Standards, Meta empirical Assessments

On the strength and the limits of the Bad Lot Argument

Alexandros Apostolidis * [1]

[1] National and Kapodistrian University of Athens – Greece

Van Fraassen formulated one of the strongest anti-realist arguments to undermine the reliability of Inferences to the Best Explanation (IBE). The usual formulation of an IBE argument is as follows:

- P1: a collection of facts, observations, or theories D.
- P2: if hypothesis hi were true, then hi would explain D.
- P3: no other hypothesis explains D as well as hi.
- C: hi is (probably) true.

Van Fraassen's argument, known as the Bad Lot Argument (BLA), is based on the fact that the set of hypotheses H from which the best explanation will be selected is finite and does not exhaust the logical space of possibilities. Thus, in every IBE argument, there may be infinitely many alternative hypotheses that have not been formulated but explain the data D just as well. Therefore, the true hypothesis may be among those not considered in the candidate explanations. In this way, it is more likely that the best explanation in IBE arguments is false, while the true explanation might be among the hypotheses not taken into account.

Many philosophers adopt a version of BLA where it is argued that the pool of candidate hypotheses used in IBE is not only small but disjointed. For example, Brad Wray argues that, with the exception of the 17th-18th century astronomical debate, scientists almost always choose between two or, more rarely, three hypotheses. Similar conclusions are found in the papers of Valeriano Iranzo, as well as in other texts, mostly by anti-realists.

The purpose of this presentation is to investigate historical examples from various scientific fields, such as ancient Greek astronomy, early 20th-century biology, and modern economic theory, where the alternative hypotheses were three or more. The aim is to support the position that the pool of candidate explanations is not as "poor" as some philosophers suggest. On the contrary, in every critical moment when either the scientific community as a whole or part of it decides in favor of a theory, the theory is not strictly selected between two competing theories but from a broader set containing a plurality of hypotheses.

*Speaker

At the same time, the strength of BLA against successive applications of IBE will be explored. Let us assume that in a scientific field, successive IBEs are required as new data continuously accumulate, but may not be adequately explained by the current best explanation. Then, the best explanation may be chosen from a set that even includes the hypotheses that have been rejected in previous stages of scientific research. If indeed the selection of the best explanation is made each time from an increasingly rich pool, this fact might argue in favor of a progressive view of science. Even if there are no guarantees that the true hypothesis was in the initial set of hypotheses H, during successive IBEs and the successive enrichments of H, the probability of the true hypothesis being found within it increases.

Keywords: Inference to the Best Explanation, Bad Lot Argument, Scientific Realism

Scientific Communication: Persuasion, Relevance, or Full Transparency?

Leon Assaad * [1]

[1] Munich Center for Mathematical Philosophy, LMU Munich – Germany

Agent-based models in social epistemology help us understand how scientists should communicate to collectively arrive at accurate beliefs about target hypotheses. Various modeling frameworks provide differing insights into the (dis)advantages of scientific communication (e.g., Zollman 2007, Hegselmann et al. 2006, O'Connor and Weatherall 2018), and thus into the optimal structure of scientific communities. However, these models rarely explicitly link to the less formal literature on argumentation.
We argue that sharing scientific results (evidence) is akin to an argumentative act. The choice of which result to present depends on the argumentative context-the epistemic goals of the interlocutors, their backgrounds, and sometimes non-epistemic intentions. The literature on argumentation examines various contexts suggesting different rules for argument deployment (Dutilh Novaes 2022): argumentation as consensus-oriented, persuasive, truth-conducive, and more.

Which argumentative context proves best for science? Which context most effectively leads groups to converge on true beliefs about a hypothesis? Here, we formulate and contrast three plausible contexts, inspired by the literature (e.g., Bright and Heesen 2023; Mercier and Sperber 2011):

1. **Persuasion**: Scientists share their most persuasive evidence to convince peers of their own positions.

2. **Relevance**: Scientists share their most relevant (i.e., diagnostic) evidence to inform peers of their most impactful findings, regardless of their position.

3. **Full Transparency**: Scientists share all their findings indiscriminately.

To analyze these scenarios, we use the NormAN framework-short for Normative Argument Exchange Across Networks (Assaad et al. 2023)-in an agent-based simulation model. Bayesian agents deliberate on whether a central hypothesis is true or false, by gathering evidence and strategically sharing their results with peers. The model captures problem difficulty: Are there enough diagnostic tests that scientists can perform to determine the truth of the hypothesis? Which gathered results are shared depends on the chosen context (i.e., most persuasive, most relevant or full sharing). While agents filter their evidence, they do not lie and interpret received evidence uniformly and correctly (i.e., according to their true likelihoods).

*Speaker

Which context proves to be the most effective? Our preliminary results suggest: it depends. Although most simulated communities perform quite well, persuasion-based contexts are the most volatile in comparison: agents may quickly converge to accurate beliefs or drift into completely wrong positions, possibly even polarizing. In contrast, relevance-based contexts generate more stable and moderately accurate beliefs in a short time, making this approach quite efficient. While full transparency ensures a perfect consensus belief as a stable state, it comes at a price: first, this consensus is much slower to achieve, and second, the consensus belief only proves to be as accurate as the evidence allows (i.e., given the difficulty of the problem).
We conclude that the optimal context depends on the epistemic goals of the community, the pragmatic aspects of the situation (e.g., urgency, cost of subpar consensus), and the nature of the evidence. Perhaps surprisingly, our results suggest that in certain situations, full transparency may be a less desirable norm than persuasion or relevance-based sharing. We connect our results back to the non-formal literature on argumentation.

Keywords: scientific communication, social epistemology, agent, based model, Bayesian epistemology

Herbaria and epistemic injustice: The legacies of local plant collectors in the history of botany

Jan Baedke * [1]

[1] Ruhr-University Bochum – Germany

Herbaria hold nearly 400 million preserved plant specimens worldwide. They are vital repositories of botanical and environmental knowledge. After decades of chronic underfunding of botanical research, recent digitization efforts are rapidly changing this situation by creating virtual herbaria that rejuvenate botany. These digital collections support studies in herbarium genomics and AI-driven evolutionary and morphological research, which are crucial for addressing conservation, biodiversity, and climate change issues. In this talk I conceptualize herbarium specimens as unique 'knowledge hubs' that encapsulate biological, environmental, historical, and socio-cultural insights. I explore these special hubs and the impact of digitization on botany through a case study on influential botanical expeditions of the Yucatan peninsula (Mexico, Guatemala and British Honduras/now Belize) since the early 20th century. This paper focuses on the role played by local collectors, like Percy H. Gentle and Elias R. Contreras, that gathered valuable botanical knowledge in these expeditions. Despite their contributions, these collectors remain largely forgotten today. By drawing on expedition reports, archival material and especially by unearthing the digital trails collected plants left in today's virtual herbaria and biodiversity collections (e.g., GBIF, SERMEC, iDigBio, Tropicos), I, first, reconstruct the contributions of these collectors to specific botanical expeditions. Then, second, this paper identifies patterns and different forms of epistemic injustice that influenced the collection, storage, and distribution of local knowledge. Third, it discusses how colonial biases still lead to information loss and epistemic blind spots in herbarium digitization today. Finally, I develop solutions to these old and new forms of epistemic invisibility of local knowers and knowledge in botany's digital era.

Keywords: Herbaria, botany, epistemic injustice, local knowledge, digitalization

*Speaker

Tractability and Model Transfer: Motivation or Barrier?

Murat Bakeev * [1]

[1] Leibniz Universität Hannover=Leibniz University Hannover – Germany

The phenomenon of model transfer has received much attention in recent philosophy of science. For instance, philosophers investigated reasons for the cross-domain use of the Lotka–Volterra model, which originated in population biology and later was also used in fields as diverse as medicine, marketing, and economics. Another example is the Ising model of ferromagnetism and its various modifications, which has been applied in neuroscience to study pattern recognition, in biology to study mutation and selection, in sociology to analyze opinion formation, and finance to better understand market crashes. Generally, the philosophical literature studying model transfer relies on the template-based account of models inspired by the research done by Paul Humphreys. On this view, model transfer is understood as the application of an existing computational or formal template, which is defined as a piece of formalism or mathematics, for example, a system of equations or an algorithm, upon which a particular model or class of models can be built. The paper explores challenges that may arise when transferring models from one scientific domain to another, based on the case of transfer of agent-based models from computer science to macroeconomics. Challenges encountered by agent-based models in macroeconomics are studied that might question the implications of the template-based account of model transfer developed by Paul Humphreys. It is argued that the Humphreys' account, according to which models become attractive for transfer due to advantages in computational tractability, does not take into account the discipline-specific contingency of tractability standards. The standards of tractability established in a discipline may differ depending on the positive or negative experience accumulated in the discipline in using various mathematical and computational techniques for constructing and solving models. As a result, the very concept of tractability in model-based science becomes path-dependent and it often differs drastically from the ability to obtain numerical solutions for a model, which Humphreys associates with computational tractability. Practicing scientists can narrow down the types of acceptable solutions that make a model tractable based on additional criteria such as transparency or stability. Specific methodological requirements for models developed in a discipline and associated with a discipline-specific interpretation of tractability may act as a barrier to the transfer of models, despite the advantages of these models in terms of the ability to provide numerical solutions. Using the example of macroeconomics, I show how a methodological standard requiring the use of structural models with invariant parameters has become a challenge for the transfer of agent-based models. Although agent-based models guaranteed computational tractability and could capture the general features of social interaction in macroeconomic systems, the domain-specific standards of macroeconomics were not fulfilled, which resulted in resistance to the transfer among many macroeconomists.

*Speaker

Keywords: model transfer, tractability, economics

On Logic and Pre-Logic

Luis F. Bartolo Alegre * [1]

[1] Ludwig-Maximilians-Universität München – Germany

Historically, logic emerged from practices of reasoning and argumentation, but these early practices didn't themselves constitute formal logic. Thinkers like Zeno and Socrates used reasoning patterns resembling methods such as *reductio ad absurdum*, yet they offered no formal account of their validity. Similarly, Parmenides advocated the invalidity of contradictions, but without a systematic theory of reasoning. These early thinkers demonstrated a basic understanding of sound reasoning, but it wasn't until Aristotle that these practices were formalised into a coherent system of logic. This doesn't imply that reasoning before Aristotle was irrational-only that the principles of validity had not yet been formalised.
In interactions between competent speakers of any language, we encounter arguments that could be classified as (a) valid according to all logical systems, (b) universally invalid all logical systems, or (c) valid according to some logical systems but not to others. Logical monists focus only on (a) and (b), leading authors like Levy-Bruhl to claim that so-called 'primitive mentality' was 'pre-logical', because it didn't conform to classical logical principles such as the law of non-contradiction. Some logical pluralists, on the other hand, argue that differences in reasoning may arise from the use of different logical frameworks within various cultures.

In this talk, I defend a *pre-logicality* thesis, but of a very different kind. I will argue that Western society before Aristotle was pre-logical-not because it was irrational, but because it lacked a systematic criterion for distinguishing valid from invalid inferences. Even today, despite Aristotle's influence, people often reason in ways that don't conform to formal logic, including logicians themselves when working outside their discipline. Levy-Bruhl's mistake was to critique the reasoning of 'primitive' societies without recognising the everyday pre-logical reasoning that persists within Western cultures.
Crucially, the distinction between logic and pre-logic I'm proposing isn't based on the assumed correctness or incorrectness of reasoning within a given culture. (There aren't enough cross-cultural psychological studies about inferential practices, and those which exist don't provide sufficient evidence on difference in such practices across cultures.) Rather, it hinges on whether or not a systematic criterion for validity exists. While a logical system could be grounded in non-classical inference patterns, it only qualifies as 'logic' if those patterns are formalised into a framework for evaluating arguments. This perspective has important implications for discussions about whether certain cultures are scientific or pre-scientific, where the focus should be on the presence of systematic methods for classifying and predicting phenomena, rather than the mere presence of knowledge.

Keywords: Aristotle, reasoning, logical pluralism, history of logic, non classical logics

*Speaker

Deeper Explorations of Wild-Type Scientific Realism

James Beebe * [1]

[1] University at Buffalo – United States

Building upon other recent work in experimental philosophy of science, we report the results of a new empirical study that examined the views of several hundred scientists about various aspects of scientific realism, including the mind-independence of the objects and phenomena studied by science, the objective truth or falsity of scientific theories, the epistemic status of scientific theories, instrumentalism, and the status of unobservables. Employing a questionnaire with greater psychometric validity, reliability, and scope than previous studies, we found that metaphysical issues concerning the mind-independence of the objects and phenomena studied by science, together with considerations from the No-Miracles Argument, were most central to scientists' thinking about self-report and indeed were significantly more central than their opinions about instrumentalism or the Pessimistic Induction. A variety of statistical analyses showed that realism or antirealism about unobservables failed by a very wide margin to fit into the overall set of ideas that constitute scientific realism as it has been discussed in recent decades. Furthermore, in some sciences, antirealism about unobservables correlated positively to a small extent with a realist attitude toward metaphysical, epistemic, and semantic aspects of scientific realism. We also observed a greater inclination to endorse scientific realism among natural scientists than among social scientists and among men than among women, and we observed an increasing tendency to endorse realism with increasing age. We supplement the latter findings with results from other studies we have performed in which men and older individuals have endorsed scientific realism and other forms of realism to a greater extent than women or younger individuals. We speculate on the competing roles that the wisdom of experience and cognitive aging might play in explaining these age-related data.

Keywords: scientific realism, experimental philosophy of science, age, gender

*Speaker

Can DNNs have epistemic authority?

Claus Beisbart * [1]

[1] Universität Bern = University of Bern = Université de Berne – Switzerland

In the last decade or so, deep neural networks (DNNs) have acquired a fantastic track record of successful applications in prediction and classification tasks within science and beyond. In many cases, it is fair to say that humans have a reason to believe that p if p is the result of a classification or prediction by a DNN. But this doesn't yet settle the question of whether DNNs can have epistemic authority, as many scientists are supposed to have. The question of this talk thus is whether and under which conditions DNNs can have epistemic authority. This question is essential, e.g., in cases where a human researcher and a DNN reach different conclusions. If the AI has epistemic authority for the human, then the latter should defer to the judgment of the AI. Ferrario et al. (2024) have recently considered a similar question, but their focus is on understanding rather than knowledge, on which I focus.
According to the canonical account of epistemic authority given by Zagzebski (2012), which is based upon work by J. Raz, the so-called preemption is crucial for epistemic authority: We do not just have reason to believe p if the epistemic authority believes p, but also to bracket our own reasons for or against believing p. Taking a DNN to be an epistemic authority is thus only justified if we have reasons to bracket our own reasons in favor or against p. One justification for doing this considered by Raz and Zagzebski is a track-record-based justification. The idea is roughly that an agent has made the experience of being epistemically better off when deferring to the DNN and not considering their own reasons.

I argue that the track-record-based justification of epistemic authority is, in principle, available for DNNs. However, this kind of justification is often not viable in practice. One reason is that the track-record-based justification needs the assumption of a well-defined reference class, which is a problem for AI because generalizability and robustness are too poorly understood. Another issue is that the track-record-based justification makes an often unrealistic independence assumption. To avoid double standards, I compare DNNs to human authorities and discuss whether the conditions are better fulfilled there.

I then move to a more general justification condition provided by Zagzebski and consider whether it can be fulfilled although the track-record-based justification fails. In particular, I consider the possibility that I recognize others as authorities because I have reasons to think that my own evidence is misguided or misinterpreted. I argue that this is not an attractive route to the epistemic authority of DNNs, either.
Overall, DNNs can, in principle, have epistemic authority, but there are significant hurdles in practice.

*Speaker

Keywords: AI, epistemic authority, justification, track record, misguided evidence

Cassirer on Object, Measurement and Technology

Spigola Benedetta * [1]

[1] Università degli studi di Torino = University of Turin – Italy

In this paper, I will focus on Ernst Cassirer's reflections on technology in order to argue that the concept of empirical data in physics is not just what is given to observation, but rather it is structured according to certain conditions for measurability and the construction of technological instruments for measurement. I will do it by taking into account Cassirer's analysis of the process of measurement. Indeed, measurement represents the most basic level of physical knowledge, i.e., the sphere of empirical data. This is what physical theories are said to refer to and is the first step in the construction of theories.
I will focus on what Cassirer calls 'statements of the results of measurements'. I will show that Cassirer's analysis of the conditions of the possibility of physical measurement allows to consider the empirically measured object as being intrinsically determined by both formal conditions (e.g. logical rules, algebraic formula and geometrical models) and material conditions (e.g. natural laws, physical principles and technological resources), with these conditions determining the context of meaning whereby the empirical data are structured and can be interpreted *qua* objects of physics. Moreover, I will show that the praxis of measurement makes it possible to grasp how technology bears on and interacts with the theoretical activity in what Cassirer calls the "act of world creation". Finally, I will focus on the interplay between measurement and technology in order to show that the latter is the channel through which physics 'gets in touch with' the natural world, this being no longer conceived as essentially separated from the structure of knowledge.

The philosophical figure of Cassirer is rather underestimated in the context of the studies on technology during the first decades of the 20th century. I think Cassirer's reflections on technology help clarify what makes technology an instrument for physical knowledge, without being *materially reductive*. Cassirer states that technology is genuinely analyzed if *technological efficacy*, rather than the entity being taken as a technological instrument, is taken into account. In this light, the instrument is understood to detect something specific in the natural world of phenomena, e.g. the electric current, and this observational-theoretical purpose is what governs, at once, the construction, use and interpretation of the output information. Therefore, the 'thinghood' of an instrument is intrinsically determined by the purpose to which it is oriented. The technological instruments is inherently linked to its being recognized as a *goal-directed entity*; this means that it must be possible to put the technological instrument to use for physical knowledge. This will allow to state that technological creativity is always structurally embedded in the theoretical activity of theory-building, namely, it is based on the capacity to frame output information against a theory of the natural world that allows us to make statements of measurement.
I will conclude by highlighting how this activity is necessarily bound to formal and material

*Speaker

conditions governing the possibility of making things and events determinably measurable and, therefore, turning them into objects of physics.

Keywords: Cassirer, measurement, technology, object, instruments.

The practical goal of scientific concepts

Birgit Benzing * [1]

[1] Leibniz Universität Hannover=Leibniz University Hannover – Germany

Concepts play important roles in scientific practice. In traditional philosophy of science, concepts were primarily considered in terms of their contribution to theory formation. The understanding of concept was rather static and focused on definitions and reference. However, many scientific concepts are complex, vague, changing – and yet, they are fruitful and make a significant contribution to research. Numerous new approaches in science research deal with these contributions regarding reasoning, experimenting, modelling and explaining. For example, concepts help to describe phenomena. They serve as tools for investigating and delimiting the object of research and forming operational definitions for measurements. How a phenomenon is conceptualized affects scientific practice and the way in which research is conducted. These analyses go hand in hand with developing more dynamic notions of what a concept is, which can, for example, explain conceptual change as rational, or explain the structure of fragmented concepts. Such accounts propose to understand the epistemic goal, i.e. what a concept is used for, as a component of the concept.

In my presentation, I will build on this development. However, I will propose that not only epistemic, but also practical goals should be considered as part of scientific concepts. I will develop my argument in four steps: I will show that scientific concepts have goals. These goals can be epistemic or practical. If the research area is characterised by societal concerns, practical goals may dominate over the epistemic goals. In these cases, the practical goal becomes part of the scientific concept. If, moreover, the research area is inter- and crossdisciplinary, the concept can have a coordinating and constitutive effect on the research area through the practical goal – even at the cost of conceptual difficulties.

I focus on a specific area of applied research that is characterised by these two features, social, political and ethical (in short: societal) concerns, and inter- and crossdisciplinary approaches. Animal welfare science stems from societal concerns on how we treat animals. It aims at developing measures to improve the treatment and the husbandry of animals. This practical goal is pursed by numerous research approaches, such as behaviour biology, veterinary science, and cognitive science. The scientific concept of animal welfare is subdivided into multiple conceptualisations that address different perspectives of the complex phenomenon and are embedded in different theories. The concept provides a common framework for research, assuring that all research aims at fulfilling the practical goal to improve animal husbandry.

Keywords: Scientific concepts, epistemic goals, practical goals, conceptual structures, applied research, animal welfare science

*Speaker

The modal-explorative understanding of Newton's experimental methodology

Daian Bica * [1], Markus Schrenk† [1]

[1] HHU Düsseldorf – Germany

In this presentation we aim to show that Newton's experimental philosophical methodology exhibits a modal-explorative nature (cf. Steinle 2006/2016, Gelfert 2016, Massimi 2019). First, we shall argue that Newton's methodology of queries can be better qualified as being explorative in a modal sense (cf. Newton 1959-1977, 1713/1999, 1717/1952), i.e., Newton provides under the label of queries various scientific models with non-actual/fictional target-system (e.g., aether models with low resistance from his "Queries" to the *Opticks*). Queries have the methodological role of suggesting further experiments and advancing theoretical points for extending the boundaries of Newton's existing knowledge. *Via* his aether models, Newton aims to find out plausible possible causal mechanisms for the production of gravitation. In this place, we shall distinguish between an *ontic* and *epistemic* sense of being a targetless model. In the ontic sense, Newton's models lacked an *ontic target*, i.e., the purported phenomenon does not putatively exist. However, aether models have an explicit *epistemic target* because Newton aims to get knowledge about how the law of gravitation is causally produced. Second, this modal-explorative methodology has immediate consequences over his metaphysics of laws of nature. Recent commentators (e.g., Biener & Schliesser 2017, Walsh 2017, Henry 2020, Hazelwood 2023) remarked that Newton's definition of active principles as laws, causes, natures, or dispositions gives rise to various conceptual tensions. However, we aim to show that the tension inherent in his conception of laws is merely *apparent,* since he is wrestling with multiple possibilities of making a theory of laws work. According to the advanced modal-explorative reading, Newton attempts to find out a plausible contender for what laws of nature *may* be. Relative to the latter reading, his definition of active principles in terms of laws and dispositions should not be taken at face value, but seen as a "never-at-rest" explorative activity.

Keywords: explorative modelling, early modern experimental philosophy, laws of nature, aether models, targetless models

*Speaker

†Corresponding author: schrenk@hhu.de

Evaluative Categories and their Blinding Effect

Simon Brausch *† [1], Sam Ducourant * ‡ [2]

[1] Max Planck Institute for the History of Science – Germany
[2] Max Planck Institute for the History of Science – Germany

Evaluative categories, such as validity and reliability, are central to the methodological debates about scientific knowledge. If a method is said to be valid, reliable, or accurate, it is assumed to be trustworthy, i.e. to be a useful and efficient tool. An invalid or unreliable method, on the contrary, should not be trusted and its results treated with caution, if at all.
In the present paper, we argue that such evaluative categories can and have been misused. They can have what we call a blinding effect. Though they do help in producing knowledge, they sometimes prevent scientists from using tools (that is, scientific methods) in the most generative way. We propose to study how the use of evaluative categories has (mis)led scientists into over-hastily accepting or rejecting a scientific method and the knowledge produced by it. We compare two historical case studies of concrete uses of these categories, in the biomedical and agricultural scientific practices, with a philosophy of science-in-practice approach.

The first case study looks at quality assessment tools for clinical trials. They emerged in the 1960s and have been widely used by meta-analysts and regulators since the 1980s. For these tools, inter-rater reliability has established itself as a category by which they must all be evaluated. But historical research shows that most and especially early tool-makers did not consider reliability as an evaluative category that their tools should meet. Instead, it was only because of how these tools were misused in meta-analytic and regulatory practices that concerns about reliability were forced upon the standardization of quality assessment. But, we claim, thinking of quality assessment tools in terms of reliability is misleading. It is not only impossible to achieve (a tension between theory and practice), but also obstructs the view of how they could be used differently.

The second case study is the notion of "welfare", as used by the early animal welfare sciences (1920's-1950's, UK). This notion first came into being in the political debate about anti-vivisection, i.e. whether animals should be used for scientific experimentation. But archives at the crossroads between scientific and activist practice show that the very use of this term was intended as a sociological weapon: it was elaborated as a means to protect science against anti-science, anti-speciesist claims. By elaborating the statistical tools that led to the 3Rs principles for humane experimentation (that is, in their very practices of validation), British scientists silenced calls for the abolition of animal experimentation, and translated "animal welfare" into a tool for optimizing animals' production and use.
Both cases taken together show how evaluative categories shape scientific practices – and how

*Speaker
†Corresponding author: sbrausch@mpiwg-berlin.mpg.de
‡Corresponding author: sam.ducourant@ens.psl.eu

this can be obstructive. In the biomedical as well as in the agricultural case, evaluative categories have not gotten us closer to what the respective practice is good for, but rather contributed to losing sight of the actual goal. This is what we call the blinding effect of evaluative categories in scientific practice.

Keywords: Evaluative Categories, Trust in Science, Evidence based Medicine, Animal Welfare Sciences

Inference to the Best (Available) Explanation in Science

Frederick Britt *† [1]

[1] Universität Bern = University of Bern = Université de Berne – Switzerland

As a form of scientific reasoning, Inference to the Best Explanation (IBE) has been a persistent conundrum in the philosophy of science. I will not go into details about what counts as an explanation, what makes one explanation better than another, or whether the fact that one hypothesis explains the evidence better than others really means that it is more likely (or less unlikely) to be true. These are key issues, but the problem I want to address is that such inferences depend on a comparison of the *available* rather than *all possible* hypotheses to explain something. It follows that possibly none of the hypotheses (that happen to be) under consideration is true. So indeed, how can scientists make such inferences?

The problem (aka the 'bad lot objection') is notorious, but it strikes me that much of the relevant literature is misguided insofar as too many settle for a philosophical workaround (often drawing on preconceived views about the nature of scientific knowledge such as externalism, pragmatism, or probabilism) rather than examining the potential implications for scientific practice. I will argue that this is a missed opportunity, especially since the objection itself is really quite nuanced. It does make it clear that the form of IBE is unreliable because it applies even if none of the available hypotheses are true to begin with, but it would be a misunderstanding to conclude that scientists should not make such inferences at all: It is just that an inference to the best explanation is not compelling unless you already believe that probably one of the available hypotheses is, in fact, true. There is a substantial problem insofar as it may be difficult to provide reasons to believe this in a given case, but as such it would seem to require scientific rather than philosophical attention after all.

What I want to do in the rest of the paper is thus to explore, first, how scientists encounter the problem in practice and then, second, whether or to what extent their reassurances might be compelling. As for the first point, I will draw on the pragmatics of explanation, arguing that scientists are liable to focus on a narrow range of potential hypotheses by *framing* (i.e., by setting the focus and background of) their research in one way or another. That may sound like a concession but, as for the second point, note that it is a matter of debate among scientists themselves: They are, of course, accountable for any background assumptions they make, and so I will argue that an inference to the best explanation is indeed unacceptable (though often stimulating) unless the relevant scientific community *agrees* on the underlying framing to begin with.

There are some caveats, but the point is that you will have to engage with the science itself to question such inferences, as scientists might have something to say about the range of hypotheses under consideration and individual biases can, at least in principle, be balanced through open

*Speaker

†Corresponding author: frederick.britt@unibe.ch

debate.

Keywords: Inference to the best explanation, bad lot objection, pragmatics of explanation, social epistemology, scientific practice

The heterodox view of Leibniz's Principle and similar quantum particles: the ambiguity problem

Maren Bräutigam * [1]

[1] University of Cologne – Germany

According to the heterodox view, similar quantum particles like, e.g., two electrons, are (almost always) qualitatively distinct, so that Leibniz's Principle of the Identity of Indiscernibles is valid (this view contrasts with the orthodox view, according to which the opposite is the case). Bigaj (2022) has argued that the worst problem for the heterodox view, whose ultimate solution is yet to be found, is the ambiguity problem, according to which individuation via properties is not unique. I argue that three different kinds of ambiguity, namely *interbasis-ambiguity*, *intrabasis-ambiguity* and *spinsymmetry-ambiguity* (which is a subcategory of the latter) should be distinguished.

Interbasis-ambiguity is due to the possibility of performing basis transformations; i.e. it involves considering (at least) two different bases. In contrast to that, intrabasis-ambiguity arises within one and the same basis (no basis transformation is needed). Spinsymmetry-ambiguity holds if the spin-part of the state under consideration is spherically symmetric. As already shown by Caulton (2015), interbasis-ambiguity can be solved relatively easily by demanding that the individuating properties are not superposed (which is typically the case for all but one set of potentially individuating properties). However, this criterion does not help with intrabasis-ambiguity: all sets of potentially individuating properties are non-superposed. For example, in entangled states such as, e.g.:

$$1 / \sqrt{2} \; ((—\uparrow\rangle 1 —\downarrow\rangle 2 \text{ - } —\downarrow\rangle 1 —\uparrow\rangle 2) \otimes (—L\rangle 1 —R\rangle 2 + —R\rangle 1 —L\rangle 2))$$

the choice is between two particles which are individuated by position properties (left and right), and two particles which are individuated by spin properties (spin up and spin down).

I propose a new criterion which can deal with interbasis-ambiguity and intrabasis-ambiguity in a uniform manner: the measurement approach. The main idea behind this approach is that individuating properties are physical; physical properties, in turn, are properties which can be indicated by measurement devices. In cases of interbasis-ambiguity, the measurement approach disqualifies superposed properties as individuating properties, because superpositions cannot be measured (if they could, there would be no measurement problem). Therefore, according to the measurement approach, the particles are unambiguously individuated by non-superposed properties. In cases on intrabasis-ambiguity, the measurement approach disqualifies spin properties as individuating properties, because the only available specifications of said spin properties ('spin up' and 'spin down') are not measurable – only spin up and spin down in a specified direction are. Therefore, according to the measurement approach, the particles are unambiguously

*Speaker

individuated by position properties. The fact that no direction of measurement can be specified stems from the fact that states which display intrabasis-ambiguity typically also display spinsymmetry-ambiguity, i.e., their spin-part is spherically symmetric. However, some cases of spinsymmetry-ambiguity – namely those which do not involve two different position properties – cannot be solved by the measurement approach. An example are electrons in the ground state of the helium atom: these electrons cannot be individuated by their position properties, since these are the same for both. Neither can they be individuated by their spin properties (according to the measurement approach), since these are not measurable. Therefore, spinsymmetry-ambiguity is the hard problem for the heterodox view.

Keywords: Leibniz's Principle of the Identity of Indiscernibles, similar quantum particles

Understanding Gender Bias in Psychiatric Diagnosis

Anke Bueter * [1]

[1] Aarhus University – Denmark

Many mental disorders have differential prevalence rates in male versus female populations. For example, boys & men are more frequently diagnosed with autism spectrum disorder, psychopathy, or attention deficit/hyperactivity disorder (ADHD). Focusing on the latter example, I will argue that:
A) There are good reasons to assume that underdiagnosis in female populations relative to male populations is due to gender stereotypes.

B) However, the prevalence rates in male populations likely reflect overdiagnosis.

C) In this context, diagnostic bias is best understood not in terms of a deviation from the truth (i.e., from the correct prevalence rate), but from proper diagnostic practices and value-laden clinical goals.

A) In recent years, ADHD diagnoses among girls – and especially adult women – have risen sharply. It is often argued that this trend corrects a historical tendency toward underdiagnosis driven by gender bias. To examine this claim, I will start by looking at how exactly gender bias can affect diagnosis. For example, diagnostic categories may medicalize gendered norms of behavior; criteria may fail to capture variations in symptom presentation across genders; or gendered perceptions of disorders may influence who gets diagnosed and treated in practice.
Utilizing this, a convincing case for gender bias in ADHD diagnosis can be made. For example, ADHD tends to manifest differently in females, with girls and women more often exhibiting inattentive symptoms rather than the hyperactivity and impulsivity typically associated with ADHD in boys. Because inattentive symptoms are less disruptive in educational settings, they are less likely to prompt psychiatric evaluation or treatment.

B) Against this background, it has been argued that the rising number of ADHD diagnoses in female populations present progress towards both gender equality and scientific objectivity. However, this optimistic conclusion is problematic insofar as there are also compelling arguments that ADHD is significantly overdiagnosed (and overmedicated) in boys. In this population, too, there has been a substantial increase in diagnoses in the last decades. Whereas this might in part be due to increased awareness, it also aligns with a softening of diagnostic criteria for ADHD, and with general trends of diagnostic inflation.
C) In summary, there is evidence of gender bias leading to an underdiagnosis of females *relative* to males. However, striving for parity in diagnosis rates between males and females is problematic, given that the higher rates in boys may reflect overdiagnosis rather than a "correct"

*Speaker

baseline. At the same time, it is difficult to conceptualize diagnostic bias here in terms of deviation from the correct prevalence rates. Epistemologically, it is unclear how to figure out where the correct line lies. Ontologically, it is unclear whether there even is a definite line to figure out, since some people reject a conceptualization of ADHD as pathological in the first place. I will argue that in such a situation, "bias" should not be understood in terms of deviation from the truth, but from value-laden clinical goals to diagnose where diagnosis brings more good than harm to patients.

Keywords: ADHD, diagnosis, gender, bias, psychiatry

A Novel Response to the Fine-Tuning Argument: The Variable Conceptualization of Constants

Maura Burke * [1]

[1] Universiteit Utrecht / Utrecht University [Utrecht] – Netherlands

In this paper, I explore the conceptual foundations of and relations between the Fine-Tuning Argument (FTA), Multiverse Theories (MVTs), inflationary cosmology, and nomological necessity to argue for a new response to the FTA: a variable conceptualization of physical constants (VCC). The FTA argues that our existence is unlikely given the logically infinite range of possible values physical constants could assume. In this sense, these constants of the Standard Model(s) seem to be 'fine-tuned', that is they take on *exactly* the values (within a very small range) that are required for a universe to have the, minimally construed, necessary conditions for life to emerge. There are three typical responses to the FTA: the Design Hypothesis (DH), which postulates the existence of a divine creator who intentionally designed the universe to be life-permitting; the Just-So-Theorems (JSTs), which argue that the initial conditions of the universe simply cannot be explained and that at some point we must accept the boundaries of our explanatory practices; and the MVTs, which postulate the existence of a(n) (infinite) multiverse in which the values of the constants under consideration randomly take on any (and all) possible values, thus rendering the existence of a life-permitting universe a logical necessity. Each of these approaches has been discussed robustly in the existing literature, therefor I will only provide a cursory overview of each before presenting my novel response to the FTA, the VCC. While this position is novel, such that it has not been explicitly detailed or defended in the existing literature, one can find tacit considerations of the position in the literature on MVTs and inflation and thus has precedent. Inflation is typically deployed as the candidate physical mechanism that could produce a(n) (infinite) multiverse to increase the legitimacy of the MVT response to the FTA. I will adopt a similar physical justification for the VCC, however, I will argue that these considerations motivate a reconceptualization of our commitment to nomological necessity and the immutable nature of constants in our metaphysics. The image that emerges from these considerations is not a multiverse, but instead one universe across which the values of constants are localized, not global, and can vary with sufficient distance. Further, I will show how this commitment is tacitly made in the existing literature and clearly define what we expect and desire from explanatory elements such as laws of nature and constants. I argue that developing a less generalized vocabulary that can robustly distinguish between governing and descriptive understandings of 'necessary' *explanans* is essential, possible, and conceptually liberating. The VCC allows us to take seriously and make sense of the FTA without resorting to the ontologically expensive MVT conjecture while also stimulating a novel stream of analysis within the philosophy of science, the consequences of which are not restricted exclusively to considerations of the FTA but extend to touch upon the role nomological necessity plays in the construction of and deliberation between theories across the sciences.

*Speaker

Keywords: Necessity, cosmology, laws, constants, explanations, fine, tuning, inflation, nomological

CANCELLED

C. R. Darwin (1809-1882), J. C. Bose (1858-1937), and the problem of plant sentience

Manjari Chakrabarty * [1]

[1] Department of philosophy and comparative religion, Visva-Bharati University – India

Plant-sentience, inarguably, is quite a thorny issue. Most plant-biologists would reject any association of sentience with plant life. However, the recent resurfacing of Darwin and Bose's arguments for plant intelligence or sentience premised upon their decades of intensive research and ingenious experiments is what motivates the present study on the very possibility of plant-sentience. We aim to address some of the foundational challenges faced by the emerging (inter-disciplinary) field of plant-sentience.
Fortunately, plant-sentience is no longer as forbidden a topic in philosophy of science as it has been some years ago. As the last decade witnessed the birth of the new discipline of plant neurobiology, the tides seem changing. Profound and provoking questions like "Are plants intelligent?", "Can plants learn?", or "Do plants behave purposefully?" are now being raised by scientists and philosophers alike. The time seems ripe for bringing the debate about plant-sentience before philosophers of science particularly interested in the scientific study of consciousness.

Attributing sentience to plants amounts to implicitly acknowledging the controversial fact that sentience can be realized in systems with a substantially different organization from human brains. Sentience may be interpreted in two different senses. In a broad sense, sentience refers to any capacity for conscious experience, or what is often called 'phenomenal consciousness'. In a narrow sense, sentience refers specifically to the capacity to have experiences that feel bad or feel good to the subject-such as experiences of pain and pleasure. The present study intends to use the term 'sentience' in an even narrower sense referring to the ability of electrical responsiveness (or mechano-sensitivity) of complex systems.

Based on foundational principle of plant neurobiology (the controversial 21st century movement integrating three historically marginalized subdisciplines, namely, plant ethology, whole plant electrophysiology and plant comparative psychology) that plants are exquisitely (mechano)-sensitive organisms and may thus be seen as "minimally cognitive", this paper unfolds as follows. Upon examining the serious and persistent disagreement about the very possibility of plant-sentience in the existing literature, the first section highlights the recent advances in plant neurobiology and plant cell-biology uncovering some surprising similarities between plant cells and the neurons of animals. The second section focuses on Darwin's and Bose's research on plant-sentience, which transcended disciplinary boundaries and challenged the (apparently) unbridgeable Aristotelian division of non-human life into *anima vegetativa* and *anima sensitive.* A special emphasis is put on the radical re-conceptualization of plants as knowledge-accumulating

*Speaker

systems that enact many of the same behaviours as do animals (despite lacking an obvious brain) and the critical role played by Darwin and Bose in instigating this paradigmatic change in how to regard plants conceptually. The paper concludes by commenting on why, despite Darwin and Bose's findings of plants having greater neuro-ethological abilities than is generally assumed and the more recent discovery of plant cells showing action potentials (APs), plant-sentience as a topic is yet to find place in plant biology textbooks.

Keywords: Charles Darwin, J. C. Bose, Plant Sentience, Plant Neurobiology

Data before purpose, and the risks of overlooking biology

Emma Cavazzoni * [1]

[1] Technical University of Munich – Germany

Data have acquired a prominent status as research components and key assets over the last two decades. This is due, on the one hand, to the emergence of technologies and institutions focused on data-intensive methods and novel forms of computational analysis (e.g., AI). On the other hand, it follows a broader societal trend towards the datafication of human life and interactions with technology, whereby the digital traces left by widespread smartphone use, digitalized services,and commercial platforms are traded and analyzed on a mass scale, providing tools to better understand and predict behaviors.
In this presentation, I argue that the drive towards large-scale production and utilization of data often results in considerable research efforts focused on such practices, frequently without critically articulating the scientific expectations underpinning data generation and the type of knowledge these data may contribute to produce. To this purpose, I begin by providing a brief overview on the status, uses and features of scientific data from the relevant literature (Leonelli 2016; Ratti 2015; Borgman 2015; Kitchin 2014; Strasser 2019; Gitelman and Jackson 2013), highlighting their role as building blocks of knowledge, and the importance of how data are used and contextualized.

Then, I consider a specific project focused on the production and analysis of biological data to exemplify the social dynamics and material/technological/institutional constraints relevant to contemporary data-intensive science. I base my reflections on 6 months of ethnographic work and collaborations with Haly.Id, a Horizon project in Northern Italy that gathers a high volume of data through innovative technologies such as drones, camera traps, and near-infrared hyperspectral imaging to limit the damages inflicted by the brown marmorated stink bug *Halyomorpha Halys*. H. halys is a highly invasive pest that feeds on fruits, seriously damaging production. While significant efforts are invested incollecting data and advancing relative technologies, the critical examination of why the data are being gathered in the first place and the potential applications thereof often end up taking a back seat.
To conclude, I address some epistemic risks of this trend, such as becoming hostage to the pressure of exploiting existing data, leading to a perverse form of conservatism (Bedessem 2021; Currie 2019) that may hinder scientific inquiry and the pursuit of novel ideas and methodologies (Falkenberg, Sigl, and Fochler 2023; Stanford 2019). Moreover, the focus on data production techniques heavily reliant on technology may undermine the incorporation of other forms of evidence into knowledge making, such as insights from relevant social actors or data from locations less suited for these methods(Bezuidenhout et al. 2017; Leonelli 2023). While the focus on technologies and data collection also creates epistemic opportunities like novel forms of expertise and cross-disciplinary collaborations, these outcomes are not necessarily linked to prioritizing

*Speaker

data technologies over data interpretation, and could arguably spring from projects where the prospective application of data technologies received as much attention and resourcing as the development of the technologies themselves.

Keywords: Data, intensive research, plant science, knowledge production, epistemic risks

Mathematics in Maimon's and Bolzano's Conceptions of Analysis as Ampliative

Idit Chikurel * [1]

[1] Ludwig-Maximilians-Universität München – Germany

Notions of analysis as ampliative, i.e., informative and certain, appear in Salomon Maimon's (1753-1800) theory of thinking (1794) and theory of invention (1795), as well as in Bernard Bolzano's (1781-1848) theory of science (1837). Both philosophers use the term 'analysis in the broader sense' to speak of analysis which is informative, while referring to Kant's conception of analysis as 'analysis in the narrower sense.' Furthermore, they illustrate how analysis is ampliative using mathematical examples. While Maimon conceives analysis in the broader sense as both propositional and as mathematical practices, Bolzano conceives analysis in the broader sense only as propositional. Moreover, Maimon's examples refer namely to Euclidean geometry whereas Bolzano's refer not only to Euclidean geometry, but also to algebra. Despite their different conceptions of analysis, I examine their works together. This is because Bolzano knew at least some of Maimon's work on analysis and because conceptions of analysis as ampliative are not very common. In fact, while Lapointe (2011) considers Bolzano to be the first philosopher to have used the term 'analysis in the broader sense,' it has been shown that Maimon already used this term several decades prior to Bolzano (Chikurel 2020). In begin my talk by explaining what ampliative analysis is, and by shortly presenting Maimon's and Bolzano's works on the topic. I then discuss the role of mathematics in thinking about analysis as ampliative, and argue that it is very likely that in some cases, mathematics was used for conceiving of analysis as ampliative, not merely for examples. I end by considering how Bolzano's claim that the relation of grounds-consequences differs from deduction helps us better understand what ampliative analysis is.

Keywords: Ampliative Analysis, Ampliative Reasoning in Mathematics, Salomon Maimon, Bernard Bolzano, Grounds, Consequences

*Speaker

Appraising AI Forerunners: The Legacy of Frank Rosenblatt

David Colaco * [1]

[1] LMU Münich – Germany

Frank Rosenblatt, developer of the perceptron, is now recognized as a forerunner of neural networks. He has been called the "elder father" of deep learning, having modeled what now can be viewed as multi-layered neural networks and speculated on learning in this kind of network. Likewise, the IEEE established the Frank Rosenblatt Award in 2004, for "outstanding contributions to biologically and linguistically motivated computational paradigms and systems." He was an active researcher until he died on his birthday, July 11th, 1971, at the age of 43 in a boating accident.

Elsewhere, Rosenblatt is depicted in less positive ways. For instance, Crevier notes that Rosenblatt's work was presented with "distinctive flourish that eventually brought the wrath of the scientific community upon him." He made strong claims about neural networks, but "he did not back them up with well-structured scientific arguments." Likewise, McCorduck reports an anonymous scientist calling him "a press agent's dream" and "a real medicine man." Indeed, Rosenblatt's work came under scrutiny from his peers, with the development and publication of Minsky and Papert's *Perceptrons* offering what were perceived to be devastating challenges to the perceptron. Following this publication, Rosenblatt was a "broken man" according to rumor.

Though these two perspectives are not in principle incompatible, they create a bit of an enigma out of Frank Rosenblatt and his legacy. Any attempt to square him as a "medicine man" with him as an "elder father" raises a handful of questions. How should we evaluate Rosenblatt's contributions to the development of neural networks? What should we make of claims made by or about him and his research in the popular press? And, perhaps most importantly, how do past and present appraisals of him relate to past and present states of AI research?

In this talk, I answer these questions by providing some context to, and correcting the record about, Rosenblatt's research and claims. In doing so, I provide a framework for: (1) evaluating the legacies of AI forerunners like Rosenblatt, and (2) evaluating appraisals of these forerunners at different times in history. Specifically, unpacking the idea that Rosenblatt was "a press agent's dream" seems timely. We currently find ourselves trying to make sense of and account for the AI hype cycle, so it is worth us taking stock of historical instances of hype and integrating them into our assessments. Further, it is worth us reflecting on how the modern ascendence of neural networks shapes how we perceive its historical antecedents. Several earlier pieces of historical work on Rosenblatt and the perceptron were written in a comparatively "bear market" for neural networks. Given that we are now in a "bull market," we should reflect on how present neural network prospects should and should not influence our appraisal of forerunners like Rosenblatt.

*Speaker

Keywords: neural networks, hype, AI, scientific communication

Fat-shaming and ”dangerous science”

Camilla Francesca Colombo * [1]

[1] Rheinisch-Westfälische Technische Hochschule Aachen University – Germany

On 1st December 2019,a woman reported to the police that she was handed in a ”fat-shaming card” at Holborn Station.That same day, hundreds of cards were distributed in all major London underground stations.The cards,signed by ”Overweight Haters LtD”, stated:”our organisation hates and resents fat people.We object to the enormous amount of food resources you consume while half of the world is starving.We disapprove your wasting of NHS resources to treat your selfish guilt.(...)”.While this message was uncommonly aggressive,”fat-shaming” is a widespread and dramatic phenomenon, especially on social media platforms.
I focus here on the ”scientific support” behind fat-shaming discourses,which is variously employed as a justification for derogatory speech and retributive policy proposals.Roughly, scientific arguments connect fatness with negative traits,which can be grouped in these three different areas:

1)psychological traits: weakness of the will, loneliness, laziness, etc.

2)poor health: increased risk of developing medical conditions(strokes, diabetes, dementia, etc).

3)social costs: increased expenditure in healthcare, negative externalities.

I investigate here the issue whether scientific evidence connecting fatness with negative traits,which grounds most fat-shaming speech in the public space,should be pursued and widely communicated.I argue that,while scientific research connecting being overweight with type2) negative traits could be legitimately pursued, though with some significant caveats,there should be a ”moral warrant” against lines of scientific inquiry relating fatness with types1) and3).To justify this conclusion,I use Philip Kitcher’s argument on whether it is morally right to pursue ”dangerous” research topics,and whether such lines of inquiry should be morally banned.My focus is whether the kind of scientific evidence used to back up fat-shaming fits Kitcher’s criteria of ”dangerous”(i.e., socially detrimental)science.

Kitcher’s argument runs as follows:

Suppose there is a group G of socially ”underprivileged”,whose lower life standards also originate from a long-standing erroneous belief that, because of some trait T,members of G are naturally inferior.A research investigation of the form ”people with T are naturally less suitable for the roles R”(null-hypothesis H)is undertaken.Moreover, these two conditions hold:

a)Epistemological asymmetry:even if H is currently officially rejected, there is a strong tendency to inflate evidential support in favour of H.

*Speaker

b)Political asymmetry:the expected utility of the research investigation is negative for members of G:even conditional to the truth of $\neg$H, little good would be done in favour of the disadvantaged.

Kitcher then evaluates the question of the permissibility of a line of inquiry as a social choice problem:if the utilitarian computation of burdens with respect to G outweighs the benefits,then the research line is socially damaging,and this justifies a moral ban.
I examine whether scientific hypotheses as "overweight people are lazier"(H1),"overweight people suffer from an increased risk of dementia" (H2),"overweight people increase expenditure in healthcare"(H3) fit the criteria above.I argue that group G of overweight people is socially underprivileged, and fat-shaming as an attitude seems to be already legitimized by beliefs that people in G have some negative traits T:making unhealthy life choices,increasing social expenditure,lacking willpower,etc.Both epistemological and political asymmetry thus hold.Therefore,this kind of scientific research is socially detrimental.

Keywords: fat, shaming, dangerous science, science policy, science communication

Arbitrary Frege Arithmetic

Ludovica Conti * [1]

[1] University of Vienna – Austria

This talk focuses on a less-known version of Abstractionism, that we will call Arbitrary Frege Arithmetic. This system aims to restore Frege's Logicist project (i.e. a logical derivation of second-order Peano Arithmetic) by adopting, as much as possible, its original tools, namely a second-order logical system augmented with a (consistent) version of Basic Law V.

My preliminary aim consists in arguing that two of the main open questions of the abstractionist projects, i.e. consistency and logicality, have a common source in what we can call the Canonical account of abstraction, that is based on the unquestioned adoption of classical logic, classical meta- semantics and of an at face value reading of the abstractionist vocabulary.

Arbitrary Frege Arithmetic arises precisely by a three-fold weakening of this account: classical first-order logic will be substituted by negative free logic, semantical indeterminacy will be spelled out in terms of arbitrary reference and the abstractionist vocabulary will be recast in order to admit a consistent revision of Basic Law V. The resulting system will recover the consistency and achieves the logicality of the abstraction.

The changes mentioned above are supported by philosophical and formal motivations that allow us to consider them completely compatible with the abstractionist spirit. In addition, the system we are presenting is triply weaker than the original Fregean one and, for this reason, it shows that Grundgesetze proposal was just too strong for the goals it was supposed to achieve.

Keywords: Arithmetic, Abstraction Principles, Arbitrariness

*Speaker

What are diseases? Results of a thought experiment

Norbert Donner-Banzhoff *† [1]

[1] Philipps Universität Marburg = Philipps University of Marburg – Germany

Nobody will doubt that pain and other symptoms are real, but what about our categories grouping individuals together as having a particular disease? Are these only constructs that help us understand and manipulate the world? Or do they correspond to a structure underlying the phenomena? These are interesting topics for current debates on realism-vs-antirealism as well as natural kinds in biology and related sciences.
I am proposing a thought experiment in which patients seen and diagnosed by their doctors are placed in demarcated areas in a large gym. Each area stands for a disease entity. Results are anything but clear cut: typical cases are rare, large variation leads to patients resembling those in other fields more than their own. Placements are often contested, and for many patients no field can be found. Boundaries vary by region or country and change frequently because of new scientific findings, but also due to political and economic influence. I document these impacts by published research evidence and my own experience in three health care systems.

The results of the experiment are put in the context of current debates about truth convergence, causal explanation, the no-miracles argument and social processes leading to the definition of disease entities. If mapped on the medieval controversy about universals, insights thus gained favour a nominalist understanding of disease entities.

Previous authors tended to approach the problem from an angle of scientific concepts and underlying mechanisms. They thus could not escape the 'realist trap'. Instead, I propose the clinical consultation of *individual patients* presenting their symptoms and asking for diagnostic assessment as the focal point of the analysis.

By this I am not saying that patients are not helped by their doctors: diagnostic categories are necessary for effective and consistent care. But on closer reflection they are "provisional formulae designed for action" (1) or even "fictions" (2), which fulfil pragmatic purposes nevertheless.

Ethical and practical reasons prevent us from conducting such an experiment in the real world. Given the ample empirical evidence underlying the results, this work is more akin to qualitative modeling than to most previous thought experiments presented in the history of philosophy. Such an approach can help evaluate and develop the conceptual apparatus of a field (3).

How diseases are understood by clinicians and the public, has practical implications (4). Re-

*Speaker
†Corresponding author: norbert@staff.uni-marburg.de

cent debates about overdiagnosis, biomedicalization, the status of contested diseases, stigma and harm, to name but a few, underline the relevance of the issues addressed by reflections on disease entities.

Literature Cited

1. Cohen H. The Nature, Methods and Purpose of Diagnosis. Lancet 1943; 241:23–5.

2. Koch R. Die ärztliche Diagnose: Beitrag zur Kenntnis des ärztlichen Denkens. Wiesbaden: J. F. Bergmann; 1920.

3. Kuhn TS. A Function for Thought Experiments. In: Kuhn TS, editor. The Essential Tension: Selected Studies in Scientific Tradition and Change. Chicago: University of Chicago Press; 1977.

4. Donner-Banzhoff N. Die ärztliche Diagnose: Erfahrung - Evidenz- Ritual. 1. Auflage 2022. Bern: Hogrefe AG; 2022.

Keywords: Medical Diagnosis, Realism, Anti, realism, Thought Experiment

Intrumentalism and gauge theories

John Dougherty *† [1]

[1] LMU Münich – Germany

Our current best theories of high-energy physics are gauge theories. Gauge theories are often said to be philosophically puzzling because they involve mathematical structures that appear to do important theoretical work and yet be "surplus" in the sense of not directly corresponding to any features of the physical world. This feature is usually thought to be at the root of most philosophical controversies over gauge theories, such as disagreements over their explanatory structure and their observational and metaphysical consequences. That is, these disagreements are often treated as different expressions of the same basic philosophical puzzle: how can a theoretical structure contribute to a theory's success without reflecting some aspect of the world?

On this standard view, gauge theories are only a problem for the scientific realist. Contemporary scientific realism is motivated by the thought that the success of scientific theories is best explained by their truth. And this thought is incompatible with surplus theoretical structure's contributing to a theory's success, for being surplus is a way of being false. But if one is an instrumentalist, then one expects no particular connection between the success of a theory and the truth of its claims about the unobservable. So if the philosophical puzzles surrounding gauge theories indeed boil down to a concern about surplus structure, then the instrumentalist need not be puzzled.

In this paper I argue that gauge theories do present a puzzle for the instrumentalist. Moreover, the instrumentalist's puzzle is prior to the realist's puzzle in the sense that the latter arises only on some answers to the former. This argument therefore has consequences for the realist and instrumentalist alike.

More concretely, I argue that the real root of philosophical puzzlement over gauge theories is a particular kind of reflexivity found in standard presentations of these theories. Textbook presentations of gauge theories emphasize to their readers that some of the theories' structures do not have any worldly correlate, a claim often reiterated in the physics literature. Philosophers have generally interpreted these claims as philosophical assertions about which features of the theory deserve our realist commitment; this leads to the standard story about what makes gauge theories puzzling, and the instrumentalist has good reasons to be indifferent towards this puzzle. But, as I argue in this paper, one can also read these claims by physicists as scientific, rather than philosophical. On this reading, such claims are reflexive specifications of the theoretical content of gauge theories, and so the instrumentalist must account for them. I indicate how this could be done and the consequences of doing so for other debates in the philosophy of gauge theory.

*Speaker

†Corresponding author: john.dougherty@lrz.uni-muenchen.de

Keywords: scientific realism, instrumentalism, gauge theories, symmetry, quantum field theory

The reliability flexibility trade-off in science

Leonard Dung * [1]

[1] Ruhr-Universität Bochum – Germany

I will argue that there is a conflict between two desiderata for science: reliability and flexibility. What are these desiderata? Science is reliable only if the results which are reported, typically, correspond to real effects. For example, if a study claims that "treatment with vitamin D for 3 years at a dose of 4000 IU per day or 10000 IU per day, compared with 400 IU per day, resulted in statistically significant lower radial BMD" (bone mineral density) (Burt et al. 2019), then such treatment with vitamin D should, typically, lead to lower radial bone mineral density. So, the paradigmatic case of unreliability is non-replicability (Machery 2020). Yet, studies can be unreliable without being irreplicable, for example if their measures are invalid or if they suffer from (unacknowledged) confounders. Reliability is necessary for trustworthiness and, more generally, for epistemic aims of science such as knowledge, truth and understanding (Niiniluoto 2024).

Science is flexible if researchers have degrees of freedom and the capacity to appropriately exploit these degrees of freedom to advance the epistemic, and perhaps non-epistemic, aims of science. Researcher degrees of freedom (RDFs) are the extent to which researchers' choices when gathering, assessing, and interpreting evidence are constrained (Stegenga 2018). Flexibility is a desideratum because, instead of applying rigid rules according to prespecified criteria, scientists should be creative problem-solvers (Hepburn and Andersen 2021; Hoyningen-Huene 2013). According to many, scientists need to and should use their RDFs, for example, in designing meta-analyses (Stegenga 2018, ch. 6) and in responding to the inductive risk of scientific hypotheses (Douglas 2000, 2009).

Pre-registration is an example of the tension between flexibility and reliability. Preregistration demands that researchers stipulate their research hypothesis, data collection plan, specific analyses, and what will be reported in the paper in advance of conducting the experiment (Wicherts et al. 2016). I will argue that pre-registration promises reliability benefits, not achievable via other means, by restricting flexibility. If true, this shows that there is sometimes a genuine trade-off between reliability and flexibility.

For example, pre-registration is supposed to rule out "HARKing" (hypothesizing after results are known). However, HARKing is not always illegitimate (Rubin 2020; Rubin and Donkin 2022). The results of an experiment may inspire a researcher to conceive of a hypothesis which is then discovered to be strongly supported by *prior* theory and evidence. In this case, it is plausible that the current experimental result constitutes an independent test of the hypothesis and that the psychological fact when the researcher came up with the hypothesis is irrelevant for its confirmation.

I will then discuss objections to my argument that there is a reliability flexibility trade-off,

*Speaker

such as the contention that contemporary transparency methods can achieve the same benefits as pre-registration (Rubin 2020) without restricting flexibility. Finally, I will discuss some considerations relevant to successfully navigating reliability flexibility trade-offs. One conclusion is that there are tentative reasons to assign higher weight to reliability than to flexibility.

Keywords: Replication, values in science, pre, registration, confirmation

Causal-Mechanistic Explanation in Astroparticle Physics

Brigitte Falkenburg*[1]

[1]Technische Universität Dortmund [Dortmund] (TU Dortmund) – Germany

Abstract

In astronomy, since Galileo, the light of the planets, moons, and stars has been metaphorically termed cosmic messengers carrying information from the celestial bodies to Earth. This metaphor still underlies the heuristic talk of cosmic messengers in current astroparticle physics (APP). Here, the cosmic messengers are quantum processes in which cosmic rays made up of subatomic particles transfer information from cosmic sources to terrestrial particle detectors. The heuristic talk of messenger particles and the underlying model of information transfer from cosmic sources to the Earth meets the criteria of Salmon (1997) for a causal-mechanistic explanation. The causal mechanism consists in the transfer of a conserved physical quantity, such as energy, from one object to another; here: from a cosmic source to a particle detector on Earth. In APP, this process is modelled in several steps, from the production of high-energy protons, e.g. due to the explosion of a supernova, to their acceleration by plasma shock waves, the secondary production of high-energy neutrinos and photons, and their propagation through the universe including their reactions during the passage through interstellar gas and magnetic fields, to their registration by a particle detector on Earth. In the absence of a unified fundamental theory of physics, these steps are described by several models with different theoretical foundations. The talk addresses the many facets of this causal-mechanistic explanation for the case of the neutrino signal measured in 2017 by the IceCube detector in Antarctica (IceCube Collab. 2018), and the signals traced back to the galactic disc (IceCube Collab. 2021, 2023). A special feature of the Ice Cube measurements is the probabilistic character of the causal-mechanistic explanation employed in it. The data analysis determines the origin of the measured signal directly from the raw data using probabilistic methods of machine learning, without reconstructing individual particle tracks, as customary in the high-energy scattering experiments of particle physics (e.g. at the LHC). To what extent and in what sense may we still speak of single causal processes here? With what justification did the physicists locate the origin of the neutrino signal in the galactic disc, tacitly interpreting the signal and its cause in terms of scientific realism? In contrast, for the 2017 neutrino signal, an important element of the interpretation was the coincidence with the high-energy gamma ray activity of the blazar TXS 0506+056, which supported the hypothesis of the following causal process: A supermassive black hole emitted two jets; the high-energy protons in the jets, which did not reach Earth due to their deflection by magnetic fields, produced neutrinos and high-energy photons that led to the coincident measurements. This hypothesis is obviously based on an entity realism regarding subatomic particles, supermassive black holes, etc. In my talk I will analyse these assumptions of scientific realism and their justification in more detail.

*Speaker

Keywords: Astroparticle physics, causal, mechanistic explanation, scientific realism.

Examining Scientific Knowledge Attribution: A Distributed Cognition Approach

Domingos Faria * [1]

[1] University of Porto – Portugal

Science, operating as a cognitive institution, is often characterized as a collective entity subject to knowledge attributions, such as "Science knows that p." This talk aims to provide additional support for Alexander Bird's (2022) argument, challenging the notion that scientific knowledge, when viewed collectively, simply stems from or supervenes on the mental states of individuals. A key underpinning of this argument lies in the distributed model of social cognition.
Within this model, cognitive labor is distributed within research teams, dividing the task of gathering evidence into subtasks assigned to different members based on their expertise. This approach results in a scenario where no single member fully comprehends each other's tasks. Consequently, scientific work may exhibit instances where communal or group evidence or knowledge diverges from individual evidence or knowledge. Specifically, cases can be drawn where a discovery contributes to scientific knowledge without any individual possessing knowledge of it. This challenges the assumption that scientific knowing necessarily supervenes on individual knowing or other mental states.

This talk not only explores the alignment of wider science with the distributed model but also aims to address central objections raised by Jennifer Lackey (2021). Lackey presented two objections: first, the assertion that embracing "social knowing" or "social knowledge" leads to significant epistemological problems, and second, the suggestion that ascribing social knowledge to collective entities can be replaced by describing these entities as being in a position to know. We contend that these objections can be countered by demonstrating that the scientific team possesses evidence and knowledge in a non-reductive manner.

Our discussion will highlight the nuanced sense in which scientific teams hold evidence and knowledge, challenging the traditional and summativist understanding of the relationship between individual and collective knowledge. By delving into the details of distributed cognition and responding to objections, we aim to contribute to a more nuanced and comprehensive understanding of how knowledge is generated and attributed in the scientific realm.

References:

Bird, A. (2022). Knowing Science. Oxford University Press.

*Speaker

Lackey, J. (2021). The Epistemology of Groups. Oxford University Press.

Keywords: Scientific Knowledge Attribution, Distributed Cognition Model, Collective Epistemology, Social Epistemology, Social Cognition

Can Objectivity Guarantee Quality? A Examination of Research Quality Control

Ann-Christin Fischer * [1]

[1] Ruhr University Bochum = Ruhr-Universität Bochum – Germany

The concept of objectivity remains a highly debated and central issue in the philosophy of science (see John, 2021; and Reiss & Sprenger, 2020 for an overview). Traditionally, objectivity refers to the application of scientific methods in a manner that minimizes bias, thereby allowing the production of reliable and verifiable knowledge. However, this understanding has been contested: it is argued that objectivity in its classical sense is elusive and that scientific inquiry is influenced by the values, assumptions, and perspectives of the investigator and the state of the society it is conducted in.

When examining the concept of objectivity usually the methods employed are the focus of attention and it is assumed that they will led to objective results. With this understanding the internal self-regulation mechanisms within the scientific community are left out. While rigorous, standardized methods are fundamental to achieving reproducibility and consistency in scientific findings, I argue that they cannot alone guarantee objective knowledge. Scientists also rely heavily on an internal system of self-control, which involves for instance peer review, debates and funding. These practices, while intended to safeguard against bias and misconduct, are themselves shaped by social and cultural values. The tension between the ideal of objectivity and the recognition of its limitations has led to various approaches for conceptualizing the aims and values of scientific inquiry, particular the importance of understanding how different goals and values-such as accuracy, precision, explanatory power, and social relevance-shape scientific practice.

In this paper, I propose a framework for research quality that takes into account these different aims and values. By introducing this framework, I seek to move beyond the binary notion of objectivity versus subjectivity and offer a more nuanced understanding of how scientific quality is understood. My framework considers not only the scientific methods but also the broader aims that underpin scientific inquiry. In doing so, I aim to highlight the importance of considering the context in which science is conducted, focusing on the special process of scientific self-control. Ultimately, this paper argues that the strong notion of scientific objectivity in the research quality control process is combined with underlying societal and personal aims that non-transparently influence the research outcome. These points will be developed by drawing on a case study on sports doping.

Keywords: values, objectivity, quality, peer review, aims

*Speaker

The Pursuitworthiness of Experiments

Enno Fischer * [1]

[1] TU Dresden – Germany

There is a growing body of philosophical literature on the pursuitworthiness of scientific theories and research programs. The pursuitworthiness of experiments, by contrast, has attracted comparatively little attention. In this talk I develop an account of the pursuitworthiness of experiments. The starting point is a decision-theoretic model including three factors: whether an experiment should be performed depends on (i) its potential epistemic gain, (ii) the ex-ante likelihood of achieving that gain, and (iii) the feasibility of the experiment. By looking at concrete examples from particle physics I will show that the model provides a better understanding of scientists' decisions to embark on an experimental project and usefully complements extant accounts of pursuitworthiness such as DiMarco and Khalifa's recent 'apocritic' model.
First, I will explore the concept of epistemic gain, by looking at the purposes of experimentation, including the testing of theories and hypotheses, the exploration of theoretical possibilities, and the development and testing of new instruments, methods, concepts, and theories. More specifically, I will compare theoretical and experimental epistemic gains. For example, in the case of theory testing one might think that the pursuitworthiness of experiments is linked to the pursuitworthiness of the theories being tested (e.g., searches for Dark Matter are pursuitworthy because of the theoretical pursuitworthiness of Dark Matter). However, there are also important mismatches between experimental and theoretical epistemic gains. For example, an experiment's potential gain can be larger than that associated with the theory because of new instrumentation and experimental technology being developed (such as new data processing routines that can be reused in other contexts).

Second, I will address considerations of the ex-ante likelihood of achieving the epistemic gain. Whether an experiment is pursuitworthy depends on how likely it is that an epistemic gain is achieved. A natural assumption is that such considerations focus on the likelihood of the primarily intended goal of the experiment (such as the discovery of a particle in high-energy physics experiments). However, often scientists perform more sophisticated risk assessments, involving considerations about potential epistemic gains that can be reaped even in case of failed experimentation.
Third, I will address the feasibility of experiments. Conceptual design reports such as the study of a potential Future Circular Collider at CERN make detailed estimations of the efforts required to achieve the collider's epistemic goals. In a decision-theoretic model such costs are to be justified by the project's gains. However, it needs to be clarified *how* this kind of justification is achieved. In foundational research cost-benefit analysis is evidently difficult because it is typically not clear how advancements in research are to be compared with material costs. Rather than in terms of a cost-benefit, I will argue, costs affect assessments of pursuitworthiness in terms of the experiment's feasibility.

*Speaker

Keywords: Pursuitworthiness, Experiment, Particle Physics

Spare the Rod: Relativity as Pure Chronometry

Samuel Fletcher *† [1,2,3]

[1] University of Minnesota - Twin Cities (U of M) – United States
[2] Munich Center for Mathematical Philosophy (MCMP) – Germany
[3] University of Oxford – United Kingdom

At the dawn of general relativity's golden age in the 1950s, Irish mathematician and physicist John Lighton Synge proposed, both in his scientific and popular writings, to reduce spatial concepts to temporal concepts. He was responding to a twofold tradition in relativity. The first part of this tradition gave operational definitions of concepts: ideal clocks for times and rigid rods for lengths. But, Synge correctly pointed out, the latter cannot exist in relativity theory, and even approximations to them will fail for large enough distances. The second part of this tradition represented both temporal and spatial concepts in the models of relativity theory with the spacetime internal, according to its sign: for one (say, positive), it represents a duration, while for the other (say, negative, respectively), it represents a length. Synge argued that the structure of the spacetime metric allows one to define (e.g.) length in terms of time, a reduction of chronogeometry to what he calls *pure chronometry*.

Synge's arresting proposal, aside from intrinsic interest for naturalistic philosophy of time, bears analogy with proposals a decade later for the "disappearance" of spacetime (or time) in theories of quantum gravity, starting with the Wheeler-DeWitt equation in 1967. Nevertheless, it has two problems. First, while Synge rejects the need for rigid rods as a primitive operationalization of length in favor of chronometry, he still retains an inadequate operationalist definition of the latter in terms of "standard clocks" and a dubious empirical assumption about the commensurability of all standard clock periods. In particular, he assumes that the "ticks" of all standard clocks that give quantitative meaning to a sequence of events with duration can be put into commensurable ratios, but this is not possible if durations can take on an irrational number of temporal units. Second, he shows how to define infinitesimal length in terms of infinitesimal durations in the context of relativity theory, but this does not extend beyond the infinitesimal nor does it affect a conceptual reduction as claimed. The mathematics of Synge's argument for reduction simply doesn't support his conclusion of conceptual reduction.

In light of these criticisms, I defend a more moderate version of Synge's thesis and sketch a mathematical argument that avoids the aforementioned criticisms. I prove that finite lengths can be reduced to the times radar signals need to probe them. It is thus not spatial *concepts* which are reduced to temporal ones, but quantitative spatial *properties* to temporal properties. The qualitative spatial concept, familiar from standard relativity theory, of two events being spacelike related to one another is retained. I suggest nevertheless that this still vindicates the spirit of Synge's reduction using concepts and mathematical tools that would have been available to him.

*Speaker
†Corresponding author: sam.fletcher@merton.ox.ac.uk

Keywords: relativity theory, space, time, reduction, history of philosophy of science

Relative Entanglement for Similar Particles

Cord Friebe * [1]

[1] University of Siegen = Universität Siegen [Siegen] – Germany

Usually, one assumes that the Hilbert space of many-particle systems has tensor-product structure (TPS), and that the indices of the factor-spaces refer to the physical particles. However, the mathematical formalism has a surprising degree of flexibility which in our case leads to a variability of structuring the composite Hilbert space: "(G)enerally when there is one entanglement ready decomposition (...) then there are many" (Earman, 2015, 330). Consequently, the notion of entanglement depends on the chosen TPS, i.e. one and same state (e.g., the singlet state) represents entanglement relative to a particular TPS but is separable relative to another. This phenomenon has been called "relative entanglement".
Concerning the indices, it implies that there are many different indices (from one TPS to another TPS), which creates the problem that reference to particles is no longer unique. This observation has been used by Muller/Leegwater (2022) to argue that, in case of similar particles, the TPS has to be preferred that carries such indices descriptively referring to particles via spatial location, say, "righty" and "lefty". In this way, individuation via spatial location could allegedly be defended against the standard view that similar particles are indistinguishable. By contrast, the same observation has been used by Fortin/Lombardi (2022) in favor of an "ontology of properties", i.e. against the view that there are particles at all. Said with Earman (2015, 327), they hold a "pragmatist" stance according to which any decomposition is as good a way as any other for defining subsystems, whereas the former are "realists", seeking for an argument to privilege a particular one.

In my talk, I will argue that this dialectical puzzle is somehow misguided. For, regarding similar particles the composite Hilbert space does *not* have a TPS. First, I will consider the contrast case of non-similar particles, with which one can show that a TPS has *physical* meaning and should carefully be distinguished from the merely mathematical fact that a composite Hilbert space has been constructed out of one-particle spaces via tensor-product. Second, I will show that the composite Hilbert space of similar particles, although it has been constructed in an analogous way, does not have a TPS, since its physical meaning is lacking in that case.

Based on these two findings, I will conclude that, in case of similar particles, the indices do not refer at all, and that we need referring terms from outside the formalism. This confirms a traditional Leibniz-strategy for the semantics (and metaphysics) of similar particles.

References

Earman, J. (2015). Some puzzles and unresolved issues about quantum

entanglement. *Erkenntnis* 80, 303–337.

*Speaker

Fortin, S. and O. Lombardi (2022). Entanglement and indistinguishability in a quantum ontology of properties. *Studies in History and Philosophy of Science* 91, 234–243.

Muller, F.A. and G. Leegwater (2022). The Case Against Factorism: On the Labels of Tensor-Factor Hilbert-Spaces of Similar Particles in Quantum Mechanics. *Journal for General Philosophy of Science* 53, 291–306.

Keywords: Philosophy of Physics, Quantum Mechanics, Entanglement, Permutation Invariance, Leibniz, Principle

Extrapolation in ecology and the problem of spatiotemporal variability

Robert Frühstückl * [1]

[1] Universität Bielefeld – Germany

In current philosophy of science, extrapolation is conceptualized as an inference from a study- to a distinct target population of interest, and the reliability of such an inference is thought to depend on the extent to which study and target are similar in relevant respects. In this talk, I will argue that this understanding needs to be extended when dealing with extrapolation in ecological systems. My argument rests on two premises:
First, extrapolation in ecology is only sometimes adequately characterized as an inference from a model to a distinct target because it often includes inferences from a small-scale, local section of an ecosystem to the large-scale system at the regional, landscape, or even biome level. Ecologists extrapolate across spatiotemporal scales when independent variables of interest cannot be directly manipulated at the extensive spatial and temporal scales at which their effect needs to be understood.

Second, extrapolation across scales adds another dimension of heterogeneity of study and target systems because ecological systems have frequent causal discontinuities. Factors that drive processes at one scale may be irrelevant at another scale. For example, interspecific competition is an important factor driving species distributions at small scales. However, it is nearly irrelevant at biome or continental levels where climatic factors are the primary driver. This implies that extrapolating ecological relationships across spatiotemporal scales can be erroneous even when study and target system appear similar in terms of causal factors and covariates. While extrapolation in philosophy has so far mainly been understood as an inference from a model to a different target, without explicit consideration of scale, the case of ecology alerts us to the fact that considering differences in scale between study and target system can be an important factor in justifying extrapolative inferences.

For this reason, I introduce the concept of *spatiotemporal variability* in this talk and contrast it with *compositional variability*. Compositional variability describes our current understanding of heterogeneity in the philosophical literature on extrapolation, which focuses on differences in the distributions of causal factors and covariates or qualitative features of mechanisms. It explains our difficulties when extrapolating from a model to a distinct target at roughly similar scales. However, when extrapolating across scales, the relevant heterogeneity to consider is not only about the internal composition of study and target but also about the variability of a system over spatial and temporal scales. Spatiotemporal variability thus describes how heterogeneity manifests across spatial or temporal extent and explains the problems of extrapolating causal relationships from small to large-scale systems.
In this talk, I will introduce and further explain the concept of spatiotemporal variability and

*Speaker

its implications for extrapolation in ecology. Throughout, I will use examples from biodiversity-ecosystem function research to illustrate my arguments. I will also show that clarifying the problem of extrapolation in ecology can broaden the philosophical debate and advance our general understanding of the multifaceted problems that arise when transferring knowledge across different systems of interest.

Keywords: Extrapolation, Ecology, Biodiversity

Scientific Understanding, Explanations and Descriptions in Deep-learning Models

Giovanni Galli * [1], Raffaele Mascella [2], Davide Fazio [2]

[1] University of Teramo – Italy
[2] University of Teramo – Italy

In the evolving landscape of artificial intelligence, the understandability and explainability of AI systems have become crucial concerns. As AI models grow increasingly complex, they often operate as "black boxes", making decisions without clear explanations. This opacity can hinder trust, accountability, and ethical compliance, particularly in critical domains such as healthcare, finance, law, and scientific research. Still, deep-learning models (DLMs) are powerful tools in order to understand phenomena, as recognised by Páez (2019), Sullivan (2022), Fleisher (2022), Jumper (2021a) and Abramson et al. (2024). Thus, on the one hand, Explainable Artificial Intelligence (XAI) aims to answer the first issue about the opacity of the DLMs, offering us ways to understand the DLMs; on the other hand, the kind of understanding gained from DLMs leads us to re-define what scientific understanding is. According to Sullivan (2022), the lack of understanding of DLMs does not limit our scientific understanding of phenomena. She argues that when we fail to achieve understanding with DLMs, it is not due to the lack of understanding of the DLMs in question but to the "link uncertainty", i.e. the lack of evidence, knowledge and understanding of how the model and its target-system are related. On the opposite side, Räz and Beisbart (2022) argue that due to the lack of understanding of DLMs, we may fail to understand a phenomenon through the use of the models. Along their line, Durán (2021) claims that what we gain from DLMs is not a genuine understanding of the phenomena and that XAI's explanations are better defined as classifications. In this paper, we first argue that a machine can explain and that XAI explanations play the role of rule-extracting. We defend the idea that if XAI explanations can capture the rules underlying the scrutinised phenomenon, they are genuinely scientific explanations. The second claim we submit is that, given understanding as a noetic-mediated state, DLMs play the role of noetic mediators for scientific understanding but present essential differences from other traditionally well-suited mediators, such as explanations, theories, and non-artificial models. Moreover, while Sullivan (2022) correctly identifies the relation between the models and their target-systems, Räz and Beisbart (2022) catch the flaw in her assumption about the kind of scientific understanding involved in these cases. There is a crucial distinction when we speak of scientific understanding with DLMs and with other models and theories. De Regt (2017) and Khalifa (2017) defend that scientific understanding (SU) is gained through explanatory information about the phenomenon under scrutiny. However, when scientists use DLMs to manipulate, navigate and understand a phenomenon, they cannot access all the explanatory information, even if they understand it. We present the case study of AlphaFold's DLMs (Galli, 2023) to propose another kind of SU, complementary to the explanatory one, namely descriptive understanding. We then present the features of descriptive and explanatory scientific understanding involved in scientific research

*Speaker

with DLMs like AlphaFold's models. In conclusion, we outline the differences between descriptive and explanatory understanding in light of the explanations provided by XAI methods.

Keywords: Deep, learning Models, Scientific Understanding, Explanations, Descriptions, XAI

Scientific Progress, the Meno Problem, and the Demarcation Problem

Daniele Garancini * [1]

[1] Paris-Lodron-Universität Salzburg = Paris-Lodron-University of Salzburg – Austria

The contemporary debate about scientific progress focuses on discussing which epistemic standings should be associated with scientific progress. Examples abound. Bird (2007) argues that scientific progress should be associated with the accumulation of knowledge. Niiniluoto (2014) replies that increases in verisimilitude can suffice. Dellsén (2016; 2021) argues that progress should be associated with the accumulation of understanding. Shan (2019) argues that progress consists of problem-defining and problem-solving activities. Even Rowbottom (2014; 2023)-who argues that this discussion should be discontinued-does not challenge the idea that questions about scientific progress boil down to a discussion about which epistemic standings should be associated with progress.
This focus on the comparative value of various epistemic standings marks a stark discontinuity with the traditional discussion about scientific progress. According to Popper (1934), Kuhn (1962), Feyerabend (1970), and Lakatos (1978), among others, a discussion about scientific progress should focus on what Popper called "the Problem of Demarcation" (1934, p.10). This is the problem-or, better, the cluster of problems-of establishing what distinguishes science from other enterprises. Despite Larry Laudan's (1983) influential argument that a study of the demarcation problem should be discontinued, philosophers of science are continuing to tackle it, especially focusing on the distinction between science and pseudoscience. Among others, Hansson (2013), Hoyningen-Huene (2013), Ladyman (2013), and Pigliucci (2013) have recently discussed this. However, this discussion is widely thought to be independent from the discussion about scientific progress. Philosophers working on scientific progress, in particular, do not discuss issues of demarcation, and indeed some of them hold that these issues are not at all philosophically significant. Dellsén, for instance, does not put any real philosophical weight on the distinction between science and non-science.
The present paper argues that the demarcation problem continues to be salient to the debate about scientific progress. I do not dispute that problems about the comparative value of epistemic standings are significant to the debate about scientific progress. However, the demarcation problem is also significant, although contemporary authors working on this-perhaps due to Laudan's influence-have come to disregard it. A satisfactory answer to the question "What is scientific progress?" must incorporate an answer to the question "What is science?"

Keywords: Scientific Progress, Demarcation Problem, Meno Problem

*Speaker

Thinking outside the box: Causation and variable-relativity

Alexander Gebharter *† [1], Barbara Osimani [1]

[1] Marche Polytechnic University – Italy

We can distinguish between more and less methodological accounts of causation. The latter are typically aiming at explicitly defining or characterising causation, while more methodological accounts focus on epistemological questions such as how causal structure can be inferred on the basis of observation and experimentation, or when a causal model correctly represents a system's underlying causal structure. For answering questions like these, one typically uses causal models built over sets of variables that are manageable for human beings or computers. They all come with conditions that, if satisfied, rule how causal relations behave within these variable sets.

In this talk, we focus on Woodward's (2003) interventionist theory of causation, causally interpreted Bayesian networks, and structural equation models. Each of these accounts comes with conditions that need to be satisfied in order for their models to correctly characterise the causal structure underlying the systems of interest. A closer look at these conditions shows that they all involve talk about causes outside or independent of the specific variable sets under consideration. This gives rise to the question of how we can or should conceptualise causation regardless or independently of these variable sets. A first and natural response might be that we can simply add all the variables involved in these conditions to the original variable sets we are interested in and that the very same conditions ruling the behaviour of causation within the original variable sets also rule the behaviour of causation in the expanded variable sets. But this only pushes the problem to the next level. Causal relations only follow the same rules within these enlarged variable sets if the respective conditions hold for these enlarged variable sets as well. But if formulated for the enlarged variable sets, these conditions involve reference to causation outside of these enlarged variable sets, and so on. We end in an infinite regress and, thus, it remains unclear what we mean when we talk about causal relations outside or independently of the specific variable sets to be modelled.

In this talk we present a conceptual account of causation that is not relativized to variable sets. It provides a way to conceptualise causation independently of specific variable sets and the conditions required for them to be able to correctly represent. This allows one to make sense of what it means that a causal model represents correctly and it also spells out in detail what causation outside one's variable sets is and how it behaves. It provides a corrective for specific causal models and a way to make sense of causation as found in the mentioned conditions of the methodological accounts. Furthermore, the non-variable-relative account of causation provides a unifying conceptual basis for all three methodological accounts discussed. They can be seen as different methodological tools targeting a single notion of causation. Finally, the proposed non-relativized account of causation also solves a specific problem for interventionism by allowing

*Speaker

†Corresponding author: alexander.gebharter@gmail.com

for a clear definition of the controversial notion of a "conceptually possible intervention".

Keywords: causal models, methodology, representation, interventions

On ‘Lacey Bias’

William Goodwin * [1]

[1] University of South Florida [Tampa] – United States

A *"bias involves a systematic departure from a norm or standard of correctness"* (Kelly 2022, p.63). Accordingly, the potential for social and political values to introduce bias into science depends upon the norms that are taken to govern the interactions of science and values. These norms, in turn, reflect the assessment of both the risks inherent in this interaction and the ideal towards which this interaction should aspire.
There are (at least) two basic sorts of risks that any idealized model of how science and values *should* interact needs to avoid. The first is the risk of ‘wishful thinking,’ which occurs when what we desire, or attach social and political value to, changes our assessment of the epistemic credentials of a scientific claim. The second risk is ‘value masking,’ which occurs when a position (claim, or policy proposal, or risk assessment) is endorsed and presented as if it is the outcome of value-neutral research when in fact specific (and generally non-universal) social and political values played a constitutive role in the arguments or research that supports the position.

Hugh Lacey (Lacey 2005) endorses the norms of ‘impartiality’ and ‘neutrality’. Impartiality means, roughly, that wishful thinking is not appropriate in the epistemic assessment of scientific claims. Neutrality means, roughly, that science shouldn’t pick favorites among viable value outlooks. Since value masking has the net effect of marshalling scientific support for one value outlook over another, working towards neutrality will also involve eliminating value masking, whether strategic or covert.

Since Lacey endorses two norms governing science and values, there is space for two different sorts of bias: biases due to departures from the norm of neutrality or from the norm of impartiality. The former sort of bias-which for lack of a better term I will refer to as ‘Lacey Bias’-is importantly different from more standard forms of bias because it results when science is organized in such a way that it systematically favors one plausible value system over another. Lacey’s own work focuses on the development and implementation of transgenic agriculture. The departures from neutrality in agricultural research result from mutually reinforcing alignments between particular scientific strategies and value outlooks. Since particular value outlooks may be socially located, particular scientific strategies can be expected to ‘resonate’ with particular social locations. The exclusive use of a particular scientific strategy when in fact other strategies–aligned with other values or social locations–are possible is a form of value masking, and a source of Lacey bias. It creates the impression that science speaks with one voice when in fact only science in value alignment with one voice is being conducted.
In this paper, I attempt to articulate a general account of Lacey bias and to demonstrate the significance of the concept. I will present plausible examples of scientific research being conducted disproportionately according to one strategy when other strategies seem plausible, and

*Speaker

then suggest plausible explanations for this in terms of resonance between a scientific strategy and a particular social location.

Keywords: bias, neutrality, impartiality, scientific strategy, social location

Artificial possibilities: neuroscientific understanding with neural networks

Bojana Grujicic * [1,2,3]

[1] Excellence Cluster Science of Intelligence, Technische Universität Berlin – Germany
[2] Berlin School of Mind and Brain, Humboldt Universität zu Berlin – Germany
[3] Max Planck School of Cognition – Germany

Deep learning permeates much of current work in cognitive neuroscience. The complexity and opacity of deep neural networks, however, invites caution and scrutiny when it comes to their potential to accelerate or hinder the accomplishment of neuroscientific goals. One of these goals is scientific understanding, often emphasised as crucial by the proponents of deep learning in neuroscience themselves (Bashivan et al., 2019; Kar & DiCarlo, 2024).
In the field of philosophy of machine learning more broadly, Sullivan (2022) argues that the opacity of deep networks itself does not prevent scientific understanding, but rather the 'link uncertainty' between models and world targets. In a different thread, Kieval & Westerblad (2024) argue that scientists have pragmatic understanding which enables them to pick the right inductive biases for research problems. Diverging from both of these camps, my interest is in the capacity of deep learning to provide modal understanding, that is - the understanding of possibilities.

According to the account of modal scientific understanding of Le Bihan (2016), modal understanding is a navigational skill through a possibility space. Since the notion of possibility remains unspecified in her discussion, I expand her account with the notions of epistemic and objective possibility (Wirling & Grüne-Yanoff, 2021).

Epistemic possibility is concerned with possibly actual states of affairs that are not ruled out by our scientific knowledge of the world. Navigation through epistemic possibility spaces enables us to gain understanding of possibilities about actual phenomena. I illustrate how this representational use of neural networks to model actual phenomena works, by focusing on the use of convolutional neural networks to model the human visual ventral stream, as well as language models to model human language acquisition. By analyzing the interaction between the extent of the idealization of the model, research goals and our background knowledge I aim to put forth concrete criteria that can help us evaluate modal understanding provided by neural networks. According to these criteria, the most prominent case of the use of convolutional neural networks to model the human ventral visual stream does not provide modal understanding because these models are too highly idealized for these research goals.
Objective possibility is about how the world could be, independently of what the world is actually like. Objective possibility claims come in diverse flavours, depending on the type of modality at issue (logical, mathematical, physical, biological). I call the use of neural networks to target counterfactual possibilities their non-representational use. How can models be good

*Speaker

guides to possibilities? I discuss the prominent credibility account (Sugden, 2000), according to which imagination is the justificatory basis for model-based inferences about possibilities, and show that its normative criteria can be fulfilled by neural networks. This is so in virtue of the runnability of these models, and the ability to constrain their architectural features in light of our background knowledge. I illustrate how neural networks help us obtain navigational skills through an objective possibility space, using case studies of compositionality, language acquisition, and generalization.

Keywords: deep neural networks, scientific understanding, modality, cognition

The Ontology of Science, Natural Kinds and Indistinguishability

Patric Harting [*†] [1]

[1] Unaffiliated – United Kingdom

Science can be broadly understood as the quest to identify natural kinds and to characterise their properties and relations. And then to use these as the basis for explaining more complex phenomena.
Kinds are considered natural if they mark real distinctions in nature, rather than human perspectives or concerns. Natural kind realists have devised a number of criteria that kinds have to meet in order for us to consider them natural.

1) Natural kinds should form kinds and membership should be unambiguous.
2) Natural kinds should permit strong inductive inference.
3) Natural kinds should participate in laws of nature.
4) Natural kinds should be useable to give a complete and unambiguous description of the world.
5) Natural kinds should form a natural hierarchy.

By meeting these criteria natural kinds can then do metaphysical work. However, philosophers don't agree on how strongly these criteria need to be met. And there is a general worry that the proposed criteria may not be sufficient to allows us to identify natural kinds unambiguously.

I develop the view that members of a natural kind must be, in principle, indistinguishable from one another. I will show how this view provides a clear answer as to how best to satisfy the above criteria. For illustration, I will draw on examples from chemistry, namely a comparison between elements and isotopes. It follows that members of a natural kind have only invariant and necessary properties, which makes them well suited to play a key role in metaphysics.
Finally, I will briefly look at some consequences of the indistinguishability view. What are some candidates for natural kind so conceived? What kinds typically considered natural might not qualify? What are the practical and ontological consequences for science?

Keywords: Ontology, Natural Kinds, Chemistry, Indistinguishability

[*]Speaker
[†]Corresponding author: patric.harting@ou.ac.uk

What Practices Cannot Tell Us. A problem for Aposteriorism About Mathematical Knowledge

Paul Hasselkuß * [1]

[1] Heinrich-Heine-University Düsseldorf – Germany

Some philosophers of mathematics have argued that mathematics "cannot be regarded as *a priori*" (Ferreirós 2016, 310). When studying how mathematicians justify some of their propositions, these often do not conform to the rigorous standards of apriorism. Arguments include that mathematical justifications are tentative and fallible (Lakatos 1976ab), and that mathematicians often rely on "quasi-empirical" justifications (Putnam 1979). These arguments are echoed in historical studies (Ferreirós 2016) and in the notion of simile-proofs (De Toffoli 2021). In the talk, I'll argue that these attempts to establish aposteriorism mistake the *a priori/a posteriori* distinction: mathematical practices cannot give us the kind of evidence required to settle the distinction.
In the first part, I'll outline the strategies employed by proponents of aposteriorism. One strategy, proposed by Kitcher (1984), involves imposing strict constraints on *a priori* knowledge. However, this approach can be dismissed as ad-hoc. Another strategy, advocated by Lakatos (1976ab) and Putnam (1979), relies on an *argument from practice* (AP). This argument claims: i) 'quasi-empirical' justifications are *a posteriori*; ii) if mathematicians establish some mathematical propositions by quasi-empirical justification, then mathematical knowledge is *a posteriori*; iii) mathematicians *do* establish some propositions by quasi-empirical justification. Thus, mathematical knowledge is *a posteriori.*

In the second part, I'll criticize AP by arguing that the second premise requires revision. To understand this, the notions of *a priori* and *a posteriori* need clarification. Both concern justifications: *a posteriori* justifications involve empirical evidence, while *a priori* justifications do not. A proposition is *a priori* iff there is an *a priori* justification for it, and *a posteriori* iff there is an *a posteriori* justification for it (cf. AP). But for some propositions, both types of justifications may exist. For example, the proposition 'x is not red and green all over at the same time' can be justified *a priori* by reflecting on the logical truth, or *a posteriori* by observing x. Despite the existence of both justifications, the claim is still considered *a priori.* This leads to the following clarification: a proposition is *a posteriori* iff there is an *a posteriori* but no *a priori* justification for it.

In the final part, I apply this clarification to AP. The success of mathematicians in establishing some propositions through quasi-empirical justification is not sufficient to conclude that mathematical knowledge is *a posteriori*, as originally claimed in the second premise. Instead, the premise should be revised: ii*) if there is a quasi-empirical (hence, *a posteriori*) justification

*Speaker

for a mathematical proposition, and there is no *a priori* justification for it, then the proposition is *a posteriori*. However, this revised argument is no longer valid because, even if a proposition is established in a quasi-empirical way, there may still be an alternative *a priori* justification. In conclusion, the aposteriority of mathematics requires not only that some propositions are established via *a posteriori* justifications but also that there is no *a priori* justification for the same propositions. Mathematical practice alone cannot provide this kind of evidence.

Keywords: mathematical knowledge, mathematical practice, a priori, a posteriori, practices, justification

Navigating Data: The Role of Citizen Contributions in Marine Biology Research

Evan Josselin * [1]

[1] Institut Jean Nicod – Ecole Normale Supérieure de Paris - ENS Paris, Centre Nationale de la Recherche Scientifique – France

Citizen science has expanded significantly since the early 2000s, yet it remains underexplored in the philosophy of science, particularly concerning the role of citizen-generated data in scientific research. Despite the millions of data points produced and stored annually on web platforms such as iNaturalist and GBIF, there is limited research examining how this data is integrated into the scientific process. To better understand how citizen science is reshaping scientific practices, it is crucial to investigate how this data is produced, stored, used, and received by the scientific community.
This paper addresses the epistemic challenges surrounding citizen science data by focusing on the marine citizen science project "Plankton Planet" and the French database "Vigie Mer", which catalogs marine citizen science projects. By analyzing these two cases, I explore a key question in marine biology projects: How do participants' perceptions of data in marine citizen science influence its production and impact on research outcomes, and what are the broader implications for knowledge production in citizen science?

Plankton Planet aims to assess marine microbiome biodiversity by engaging sailors, ocean racing professionals, and the French Navy. Participants, known as "planktonauts," collect plankton data using the "Plankto-Kit," an onboard mini-laboratory that generates ecological, morphological, and genetic data. This data is gathered by vessels not traditionally used for scientific research. The Vigie Mer database offers a valuable resource for understanding the role of data in marine citizen science projects, shedding light on the technologies employed, the types of data produced, how this data is validated, and how they are disseminated through media platforms.

A comparative analysis of Plankton Planet and the Vigie Mer database highlights two key findings. Firstly, the definition of "data" in biological oceanography varies depending on the role of the actor. For example, scientists may prioritize metadata, while participants often value the visual aspects, desiring to "see" the marine microbiome during collection. Secondly, participants perceive data generation in marine citizen science as essential for advancing biological oceanography and protecting the oceans from climate change. They expect their contributions to inform scientific knowledge and public policy, and they anticipate feedback on their involvement. Meanwhile, scientists may focus on testing new instruments, building databases for other researchers, or raising awareness of the marine microbiome. These differing perceptions are not necessarily in conflict, but misalignment between the expectations of participants and the project's goals could lead to tensions. Understanding these dynamics is crucial for ensuring the success of citizen science initiatives.

*Speaker

By examining Plankton Planet and the Vigie Mer database, this paper aims to clarify how scientific data is produced in marine citizen science, the role this data plays within these projects, and how various actors interpret the data's meaning based on their perspectives of the project's goals. Ultimately, this analysis contributes to a deeper understanding of the relationship between citizen science and knowledge production in marine biology.

Keywords: citizen science, marine biology, data production, perception of scientificdata

Causal Loops and the Reduction of Causation

Jens Jäger * [1]

[1] New York University – United States

One of Hume's definitions of causation reduces it to counterfactual dependence: according to Hume, a cause is nothing but "an object followed by another ... where, if the first object had not been, the second would never have existed". Modern attempts to reduce causation to counterfactuals have long abandoned one half of Hume's proposal: it's nearly universally accepted that counterfactual dependence isn't necessary for causation-sometimes A is a cause of B even though B would still be if A wasn't. But the other half has received much less criticism: it is still often maintained that counterfactual dependence is sufficient for causation-that A causes B provided that B would not be if A wasn't. This essay has two goals. First, I argue that even weak versions of this sufficiency claim are false, based on a case involving causal loops. As a result, any counterfactualist reduction of causation entailing the sufficiency claim should be rejected-this includes the seminal accounts of Lewis and of Halpern and Pearl. Now, as it happens, some other structural equation accounts don't entail the sufficiency claim. However, I argue second, those accounts still don't satisfactorily handle causal loop scenarios, because the reductive accounts of structural equations they rely on misfire in the presence of causal loops. I'll finish with some remarks about the way forward.

Keywords: causation, counterfactuals, causal loops, reduction

*Speaker

Against standardising mean differences

Ina Jäntgen *† [1,2]

[1] University of Cambridge – United Kingdom
[2] Munich Center for Mathematical Philosophy – Germany

In applied areas of science such as education research or development economics, researchers often report the effect sizes of tested interventions to inform decision-makers. For continuous outcome variables (such as reading test scores), researchers usually measure an intervention's effect size using the *mean difference* – measuring how much, on average, the intervention increased the measured outcome – or using the *standardised mean difference* – measuring by how much of a standard deviation the intervention, on average, increased the measured outcome. Both of these effect size measures quantify how effective an intervention is, but they provide decision-makers with different information.
Should scientists report the mean difference, the standardised mean difference or both to decision-makers? Various concerns bear on which answer is correct. In line with recent work on effect sizes for binary outcome variables (e.g., Sprenger and Stegenga 2017), we focus on just one: When do standardised mean differences and when do mean differences inform agents sufficiently that they can identify which intervention is best?

This question gets bite when we consider the widespread uncertainty concerning measurement that besets much social scientific research (e.g., Rosnow and Rosenthal 2009). Researchers are often uncertain about a) how the measures used in various studies relate to each other (e.g., how different reading scores compare to each other) and b) whether these measures even measure the same property of interest (e.g., whether these scores all measure the *same* reading ability). The measurement uncertainty described in (a) and (b) makes interpreting and comparing mean differences difficult; researchers cannot say how big the change in a property that a mean difference represents is. To deal with this problem, researchers often follow the advice of textbooks (e.g., Grissom and Kim 2011): *standardise* the mean differences. Standardised mean differences can be compared and interpreted with respect to a property of interest, even given measurement uncertainty, or so it is claimed.

However, as I will argue, to facilitate rational decision-making, researchers should *not* standardise the mean differences when facing measurement uncertainty. Even given measurement uncertainty, mean differences inform decision-makers sufficiently to choose the best intervention exactly when mean differences do so absent such uncertainty. By contrast, standardised mean differences do not sufficiently inform decision-makers to choose the best intervention, either given or absent measurement uncertainty. Hence, mean differences, not standardised mean differences, facilitate good evidence-based decision-making – even in light of measurement uncertainty.

To draw these conclusions, I provide a normative decision model for choices involving mea-

*Speaker
†Corresponding author: ina.jantgen@gmail.com

surement uncertainty. Based on this model, I propose 1) a criterion for choosing the best intervention given measurement uncertainty and 2) a criterion for when an effect size provides sufficient information to make such a choice. Both criteria will make commitments on how to rationally respond to measurement uncertainty, which other philosophers may challenge. My adjacent aim is to set the ground for such debate.
I conclude with implications for the widespread practice of standardising mean differences in applied science. Overall, this paper adds normative considerations to methodologists' debates on standardised effect sizes (e.g., Baguley 2009; Cummings 2011).

Keywords: Effect Sizes, Standardised Mean Differences, Evidence, Based Decision, Making, Expected Utility Theory, Measurement

On the Epistemic Roles of the Individualized Niche Concept in Ecology, Behavioral and Evolutionary Biology

Marie I. Kaiser * [1], Katie Morrow

[1] Universität Bielefeld – Germany

This paper contributes to the debate about scientific concepts and how they are used and evaluated in scientific practice (e.g., Brigandt 2010, 2020; Feest & Steinle 2012; Haueis 2021; Novick 2023). We present a case study that shows how a niche concept guides ongoing biological research practice. Utilizing results of a qualitative empirical study conducted within an interdisciplinary biological research center, we argue that the concept of an individualized niche plays four fruitful and underappreciated epistemic roles in contemporary biology.
First, we show that the individualized niche concept shapes the research agenda of the research center on the level of the research questions being addressed, the phenomena being studied, the study design and data analysis. We argue that it does so by drawing research attention to individual variation and to individual-environment interactions. Second, we show how the individualized niche concept (and closely related concepts) facilitate explaining core phenomena related to inter-individual differences, such as how individual differences arise and change over time and which ecological and evolutionary consequences they have. Third, another epistemic role that we identify is that the individualized niche concept helps with managing individual-level causal complexity by providing clear criteria for selecting important causal factors and ignoring others. Forth, we reveal how the individualized niche concept promotes integrating local knowledge from ecology, evolutionary biology, behavioral biology and other biological fields.

Our analysis shows that some philosophers and biologists have been too quick to argue that the niche concept is superfluous (McInerny and Etienne 2012; Angilletta et al. 2019; Justus 2019; Wakil and Justus 2022). Philosophical analyses of scientific concepts should not be limited to the practices of theorizing and modeling, but rather take into account the variety of epistemic practices and roles that scientific concepts play. For instance, we advocate for centering the rich features of empirical practices in biology and roles of concepts related to driving research agenda formation and interdisciplinary synthesis.

References

Angilletta MJ, Sears MW, Levy O, Youngblood JP, Van-denBrooks JM (2019): "Fundamental Flaws with the Fundamental Niche." Integrative and Comparative Biology 59 (4): 1038-48.

Brigandt I (2010): The Epistemic Goal of a Concept: Accounting for the Rationality of Semantic Change and Variation. Synthese 177: 19-40.

*Speaker

Brigandt I (2020): How Are Biology Concepts Used and Transformed? In Kampourakis K, Uller T (eds.): Philosophy of Science for Biologists. Cambridge University Press. pp. 79-101.

Feest U, Steinle F (eds.) (2012): Scientific Concepts and Investigative Practice. Berlin: de Gruyter.

Haueis P (2021): A generalized patchwork approach to scientific concepts. British Journal for the Philosophy of Science. https://doi.org/10.1086/716179

Justus J (2019): Ecological Theory and the Superfluous Niche. Philosophical Topics 47 (1):105-123.

McInerny GJ, Etienne RS (2012): "Ditch the Niche – Is the Niche a Useful Concept in Ecology or Species Distribution Modelling?" Journal of Biogeography 39 (12): 2096-2102.

Novick R (2023): The neutral theory of conceptual complexity. Philosophy of Science 90 (5): 1021-1030.

Wakil S, Justus J (2022): The 'niche' in niche-based theorizing: much ado about nothing. Biology and Philosophy 37 (2): 1-21.

Keywords: Scientific concepts, epistemic roles, individualized niche, explanation, integration, research agenda, causal selection

Conceptual Disruptions and the Proper Roles for Machine Learning in Scientific Discovery

Donal Khosrowi *† [1]

[1] Leibniz University Hannover – Germany

Machine learning (ML) systems play increasingly important roles in scientific discovery, causing significant conceptual disruptions (Löhr 2023): central concepts we use to understand and structure scientific pursuits come under pressure. This paper focuses on the concept of '*researcher*': what does it mean to be a researcher? What abilities are necessary for this role? Do emerging ML systems already partly assume this role? What are good divisions of labor between humans and machines? Addressing such foundational questions becomes urgent in light of recent efforts to build *predictive discovery systems* that predict upcoming scientific discoveries and suggest hypotheses and experiments (Krenn et al. 2023). By outsourcing key activities relating to science's self-governance that are central to the role of *researcher*, these advances raise three major concerns about whether machines can and should perform these roles. First, predictive discovery systems raise concerns about *performativity:* predicting discoveries that humans are likely to make can change the very research pursuits that they will ultimately undertake; e.g. by following the predictions, or strategically eschewing them. Relatedly, when *recommending* novel research pursuits, this raises questions about the underlying values by which recommendations are made, e.g. recommending pursuits that are epistemically valuable as judged by humans, or pursuits that are valuable by some different standard. Second, predictive discovery system raise concerns about *autonomy* in regard to determining what projects to pursue. ML systems that partly automate such choices threaten to change the roles that human researchers play, e.g. by incentivizing them to pursue projects judged promising by potentially unknown standards and thereby ceding their autonomy in ways that intensify existing pressures exerted by the incentive structures of science. We may ask, then, whether implementing ML systems at core self-governance stages of science may change the roles of human researchers towards more mechanical, execution-level roles that undermine autonomy and epistemic agency. Equally, we may ask whether this is a bad thing or perhaps a good thing: it is unclear whether humans are uniquely good at determining promising research agendas. Both issues bear on the larger, familiar question what *science*, as an enterprise, should aim at. If machines are better at producing new knowledge, and the aim of science is knowledge production simpliciter, then machines perhaps should, as much as possible, play the role of goal-setting *researcher*. If the aim of science, as per Kitcher (2011), is to produce truth that is significant as judged by human interests and values, and these values are not fixed but rather discovered as science progresses, then humans are, trivially, uniquely situated to be researchers and machines should not encroach on this role. Third, predictive discovery systems raise concerns about novelty,

*Speaker

†Corresponding author: donal.khosrowi@philos.uni-hannover.de

creativity and conservatism: they may be substantially restricted in how novel the questions, hypotheses, and suggestions they offer can be. In sum, predictive discovery systems have the capacity to significantly intervene on the trajectories of scientific discovery. It is important to study questions about these impacts to get a firmer grasp on whether machines can, and should, assume roles distinctive of researchers.

Keywords: machine learning, artificial intelligence, scientific discovery, automated discovery, values in science

The two problems of physical computation

Matej Kohar * [1]

[1] Technical University of Berlin / Technische Universität Berlin – Germany

In this paper I identify two strands of philosophical literature about physical computation. The first stems from philosophy of cognition and attempts to elucidate the explanatory practices in cognitive science. The problem of physical computation here takes the form: Under what conditions does a physical system S realise (implement) an abstract computation C? (Chalmers, 2011; Piccinini, 2015; Shagrir, 2022) I refer to this issue as Problem-PhilCog.
The second strand of literature about physical computation is connected with philosophy of computer science and computer engineering (e.g., Horsman et al., 2014; Anderson, 2022). In this literature, computing is paradigmatically an activity performed by human computer users. The problem is to distinguish computers from other physical systems. It can be formulated as: (Problem-PhilCS) Under what conditions can a physical system S be used to perform a computation C by a user U?

I argue that these two problems are irreducibly distinct. I consider the minimal conditions a system must fulfil to count as a computing under each formulation. For Problem-PhilCog, the minimal viable conditions are given by the causal/counterfactual mapping account of computation (Chalmers, 1994). For Problem-PhilCS, the minimal viable conditions are a set of use-criteria I abstract from the literature: the system's user must be able to reliably and efficiently set it up for computation, make it transition to an end-state and read out the computation's result from this end-state. I then show that there are paradigmatic examples of computing systems in the sense of Problem-PhilCog which do not satisfy the minimal requirements from the Problem-PhilCS literature (e.g., brains) and vice versa (e.g., abaci).

Finally, I show that conflation of these two problems can motivate implausible views on the nature of physical computation. Lack of awareness of one of the problems can have similar consequences. Thus, distinguishing the problems and ensuring that proposed solutions keep within their proper scope is of utmost importance.

For example, Anderson's (2022) contention that unless computers are actively being used for computation, they do not compute, hinges on the idea that all physical computation is connected to use. But since there is another sense of computation, which does not require use, we can make sense of how computers can be seen to compute even if they are not actively directed by users.
Another example is Horsman et al.'s (2014) view on the nature of "computational entities", i.e. users. Horsman et al. consider users necessary for computation, because they view computation as dependent on representing mathematical entities. Users are needed to establish representational relationships between states of the computer and mathematical objects. But Horsman et al. also argue that such users need not have intentions to compute and need not be conscious.

*Speaker

This appears to be an attempt to broaden the scope of the account to cases of natural computing systems. However, this manoeuvre makes the process of establishing representational relations mysterious. It is preferable to recognise that Horsman et al.'s account is limited to Problem-PhilCS, and natural computing systems are out of scope.

Keywords: physical computation, computer science, philosophy of cognition, computers, neural computation

Visual Thinking – Representing and Understanding Science in Times of Crisis

Silke Körber * [1,2]

[1] Alumna, Humboldt University – Germany
[2] University of Vienna – Austria

Science communication plays an important role in modern democracies. In particular, research on climate change and pandemics are areas in which experts from different fields collaborate according to established scientific values and rules. At the same time, in both fields, communication with the general public is a key element, as with decision-makers and those in positions of power within society. Scientific representations and corresponding epistemic techniques enable us to draw conclusions or function as part of arguments. They are meant to help us describe and conceptualize the world and cope with a complex and "messy" reality. The question of how this can be done without simplifying and distorting information was already addressed by the philosopher and member of the Vienna Circle Otto Neurath. Together with the British philosopher L. Susan Stebbing, he was concerned with the promotion of a "scientific humanism" through "clear thinking" and the use of verbal and visual symbols. Furthermore, external factors such as accessibility, choice of medium and material aspects played a role. Both saw the dissemination of the latest scientific knowledge as a way to further the cognitive and emotional development of human beings as social entities, using formalised procedures and standards for the implementation of information.

However, the demands on science, scientific experts and their role and responsibility in society have changed considerably. This is particularly important as the credibility of experts and actors in the public discourse is called into question or general trust in the epistemic environment decreases. This can also be seen in sets of problems that have found their way into current discourse as the so-called "usability gap" or "representation risks".

In this paper, I explore how scientific representations at the intersection of science and art, in the form of close text-image combinations, enable us to deal with information and support a "scientific attitude" or at least potentially promote necessary changes. Against this background, I am interested in a dissemination that is not only suitable for a specific context and purpose, but is also transparent and intersubjectively comprehensible. Within this I explore the associated demands on epistemic environments, experts and their tasks – as well as their limitations. I will consider some case studies and use an approach that combines historic accounts of knowledge transfer and recent philosophical research on representation, e.g. modelling in science, in an attempt to understand how information needs to be organised to enable understanding.

*Speaker

Keywords: Philosophy and History of Science, Scientific Humanism, Scientific Representation, Values in Science, Expert Knowledge, L. Susan Stebbing, Otto Neurath

Discovering Maximum Entropy Knowledge

Juergen Landes * [1]

[1] Ludwig Maximilian University [Munich] = Ludwig Maximilians Universität München – Germany

In this talk, I investigate how to determine a rational agent's evidential constraints given her beliefs. Rationality is here construed as adherence to a principle of entropy maximisation as advocated by of objective Bayesianism. Such an agent first formulates the decision problem and then observes and/or intervenes to learn about the environment. Information acquired about the environment can come in all sorts of kinds such as statistical data from observational and/or intervention studies, testimony, observations and sensory inputs. Given this body of evidence the agent infers via all means at her disposal a set of probability functions compatible with the evidence, **E**. The agent then uses a Maximum Entropy Principle to pick out a probability function P∈**E**.
Determining evidential constraints given beliefs is know as "knowledge discovery" in computer science. I show that knowledge discovery succeeds in objective Bayesianism (even when entropy maximisation is generalised to maximisation of a Bregman divergence), if there is only one single linear evidential constraint but already fails for multiple linear constraints. I furthermore give some results for "discovering knowledge" for objective Bayesian inductive logic and a further variant of objective Bayesianism arising from a different explication of entropy maximisation.

Surprisingly, knowledge discovery for the standard Bayesian approach (subjective Bayesianism) fails. We can, in general, not discover the prior from the posterior; conditionalisation and Jeffrey updating erase too much of the prior that makes unearthing the prior impossible.

I will conclude my talk with a discussion of whether successful knowledge discovery is desirable. To make a long story short: it depends.

Keywords: Bayesianism, entropy, knowledge discovery

*Speaker

Epistemic Intimidation: How Political Interests Compromise Scientific Research

Anna Leuschner * , Manuela Fernández Pinto [1]

[1] Universidad de los Andes [Bogota] – Colombia

In recent years, there has been a strong increase in attacks against scientists, as various studies in the last couple of years have shown. This affects science across a broad range of research fields, including, for example, climate science, environmental toxicology, virology (since the pandemic), but also certain fields in social sciences, such as gender and race studies. This increase of attacks comes along with a rise of right-wing populist parties in democratic countries, fueling a general hostility towards science. In this talk, we will discuss this phenomenon, which we wish to call "epistemic intimidation," and how it affects science (and society). We will (a) introduce the concept of epistemic intimidation, (b) discuss the strategies used in these attacks, and (c) illuminate the epistemic consequences.
For (a), we propose the following definition:

Epistemic intimidation refers to attacks that specifically harm someone in their capacity as an epistemic agent.

We will then focus on a particular case of epistemic intimidation, namely the *systematic epistemic intimidation of scientists*, which is orchestrated by sufficiently powerful political and economic groups. We point out that what makes a group "sufficiently powerful" cannot be analytically determined but needs empirical investigation; decisive factors are, for instance, access to extensive financial resources or political influence via far-reaching connections to social elites. In cases of systematic epistemic intimidation of scientists, the aggressors are part of larger networks that enable them to manipulate the public debate and thus exert considerable influence on economic and political decisions. Given existing power relations, these cases significantly contribute to an atmosphere, in which scientists tend to become intimidated as epistemic agents, which can affect scientific practices in their research fields.

To examine (b), we will first briefly revisit the history of the intimidation of scientists: In the 1950s, the tobacco industry successfully manipulated the public opinion about the harmfulness of smoking by using a number of tactics, including attacks against scientists. Many of these tactics employed in the so-called "Tobacco Strategy" were adopted subsequently by other industries and interest groups and are still used today. Some of these tactics, such as "the manufacture of dissent", are used to undermine public trust in science, leading to an atmosphere in which scientists are attacked more frequently and more easily. Furthermore, strategic attacks also include specific tactics to silence scientists. For example, they target the scientists' integrity by claiming that they are biased, or they target their competence by appealing to sexist or racist stereotypes.

*Speaker

Regarding (c), we build on virtue epistemology and argue that epistemic intimidation undermines epistemic virtues that are particularly crucial in science, such as intellectual courage, open-mindedness, integrity, curiosity, intellectual resilience, and perseverance. This is not only problematic on the individual level, but can accumulate and lead to greater damage in scientific knowledge production. Using climate science as a case study, we explain how systematic intimidation has distorted the generation of knowledge about the effects of climate change.

Keywords: Attacks against Science, Hostility to Science, Climate Science, Virtue Epistemology

Symmetries and Powers-Based BSA

Vassilis Livanios * [1]

[1] University of Cyprus [Nicosia] – Cyprus

Recently some philosophers (Demarest 2017; Kimpton-Nye 2017, 2021, 2023; Williams 2019) have proposed that laws are the axioms of the systematisation of *powerful property* distributions that achieves the best balance among various theoretical virtues. According to its defenders, one of the main motivations for embracing this account (let me call it PBSA) is that it has several comparative advantages over Alexander Bird's (2007) powers-based account of laws. In this talk I focus on Kimpton-Nye's arguments that PBSA can meet problems that Bird's account faces to accommodate the fundamental symmetries.
Power theorists think that the sole source of natural modality is powers but, in contrast to that, symmetries seem to posit *constraints* upon the 'behaviour' of properties beyond those posited by their dispositional nature. Kimpton-Nye (2023) argues that symmetries are especially problematic for Bird's dispositional essentialism in virtue of their '*global*'nature, which means that that there is not plausibly any one property from which they 'flow'. He argues that PBSA is immune to this problem because it "rejects the idea that laws are so "local" as to have their metaphysical source in particular properties considered in isolation. Instead, laws are features of an integrated description of all possible distributions of all powers." (ibid., 328). Given that, Kimpton-Nye suggests that we should treat symmetries as BSA, first-order, 'global' laws.
In this talk I examine the prospects of that suggestion. The most plausible way to justify an affirmative answer to the question whether symmetries can be justifiably regarded as BSA laws is to show that symmetries are about events, not laws (and so, they cannot be meta-laws). I discern three possible routes to such a conclusion. First, the appeal to Wigner's authority, second, Hicks (2019) point that physicists often assume the global symmetries when they formulate new theories with different laws that replace the old ones, and third, Hicks' another point that, since symmetry principles inform us about either which physical systems are empirically indistinguishable or which systems are qualitatively identical, they seem to provide constraints on first-order facts. In this talk, I examine the relevant arguments and find them unpersuasive. Moreover, I give two arguments *against* the view that symmetries can be plausibly regarded as PBSA laws. The first argument is, in a nutshell, that embracing the view that symmetries are BSA laws may naturally lead to a kind of explanatory circularity reminiscent of Jaag's (2014) circularity objection to Bird's account of laws. The second argument is based on the observation that, in the context of the phenomenon of Spontaneous Symmetry Breaking, physicists invoke a hidden symmetry that is not manifest in nature to retain the symmetry of the Lagrangian of the system in the face of nonsymmetrical solutions. I argue that this aspect of the scientific practice does not seem to fit the view that symmetries are the results of the implementation of the PBSA procedure to the actual or possible distribution of properties instances.

*Speaker

Keywords: symmetries, powers, BSA

Reclaiming science's role in public policy: addressing data misrepresentation and silos

Louis Longin * [1]

[1] Ludwig Maximilian University [Munich] = Ludwig Maximilians Universität München – Germany

Policies for the public good should be supported by data. Governments should use data to make informed policies for the benefit of as many of their citizens as possible: seat belts become mandatory after they have been shown to reduce traffic deaths; anti-smoking images were put on smoking products after smoking has been linked to an increase in health risks. Science uncovered those connections and has been pivotal in providing governments with robust data to make informed policies decisions for the public good.

But the ground is changing. Now, big tech companies possess an unprecedented access to personal data and have used their power so far mostly for profit rather than public good. Cambridge Analytica and LLMs are just recent examples. So far, in attempts to protect user privacy, regulators have issued consumer protection regulations only in reaction to big tech companies, leading to a growing disconnect from the needs of their citizens. Science needs to step in. It needs to reclaim its original role as a trusted connector/bridge between individual citizens and their government.

In this paper, I will provide a perspective for how this might be done. I will focus on tackling the growing issue of data handling and sharing, which have been the driving force behind current regulations and privacy concerns. Two issues need to be addressed. One is the growing risk for data misrepresentation arising from the risk of over- and under-sharing of personal data on an individual level. The second is the prevalence of large-scale systematic data gaps, also known as data silos, that emerge from isolating the data use to the data-collecting parties. Addressing both are important steps towards regaining a strong position of science in shaping public policies.

Under-sharing occurs when data is gathered, often unknowingly, only from a partial population but is expected to represent the whole population, resulting in a skewed, non-representative societal perspective. Over-sharing captures cases of passive, indiscriminate data sharing, where extensive data is given away for digital access or small rewards. Examples include social media use and 'smart' home appliances. Conclusions from both under- and over-sharing are often unreliable and if used as the foundation for policy decisions can result in systematic discrimination and biases.

Another way to counteract data misrepresentation is aligning previously isolated data. Only aligned and connected data sets like in the Human Brain Project can answer large-scale social questions like how does the brain process information or what do elderly citizens think of new energy policies? Here, often not 'more data' is needed but rather a robust way of matching data

*Speaker

with the human mind. While Open Science provides a framework for making science accessible, Science needs to reconsider how to facilitate large-scale and efficient data sharing shared across communities and data holders.

Keywords: data privacy, data sharing, open science, citizen engagement, public policy

Do ML models in science function as toy models?

Luis Lopez * [1]

[1] Leibniz Universität Hannover=Leibniz University Hannover – Germany

In recent philosophical debates about machine learning (ML) models in science, there is an emerging view that these models are or function as highly idealized models (Fleisher 2022, Duede 2023, Sullivan 2023 & 2024). Notably, Sullivan (2023) argues that ML models function as (highly idealized) toy models and can, therefore, provide scientific understanding despite being dissimilar to their targets. According to Sullivan, both toy models and ML models can do this through 'interpretative representation' (Nguyen 2020) or 'explorative epistemic idealization.' (The names given to these two accounts are mine.) While Sullivan's view highlights how ML models, as toy models, offer understanding of their targets despite being significantly dissimilar to them, this focus on dissimilarity overlooks more fundamental aspects of ML modeling-such as the role of model-target approximation rather than idealization. Current debates thus lack an exploration of how ML models may enable understanding through approximation and model-target similarity. This contribution seeks to provide a complementary view to Sullivan's thesis, emphasizing model-target similarity and approximation over dissimilarity and idealization. The goal is to investigate whether and to what extent how-possibly explanations derived from ML models are connected to the real-world features these models are supposed to track and to provide a clearer understanding of ML models' epistemic role in scientific practice. In order to do this, I reconstruct Sullivan's arguments, examining the assumptions behind the analogy between ML models and toy models, and challenge the premise that dissimilarity is key to ML models' epistemic contribution. I develop an alternative framework, grounded in key differences between approximation and idealization, to explain how ML models can provide scientific understanding of their targets. As a result, I argue that ML models should be similar to their targets to provide understanding and introduce the Approximation Failure Hypothesis (AFH) as an alternative to Sullivan's Idealization Failure Hypothesis (IFH). The idea is that, unlike toy models, ML models aim for approximate similarity with their targets, offering epistemic value through learned patterns in data, which are hopefully aligned with real-world phenomena. In conclusion, while Sullivan's idealization framework captures some important aspects of the ML modeling pipeline, the approximation-based account provides a more fundamental understanding of the role of ML models in science, or so I argue. More specifically, the AFH better addresses the representational capacities of ML models by focusing on their approximation to real-world features, offering how-possibly explanations grounded in those approximations rather than idealizations. This analysis invites further exploration of the role of approximation in scientific models beyond ML, encouraging philosophers of science to refine the conceptual distinction between idealization and approximation in various modeling practices (especially in those that integrate ML with traditional modeling), with implications for scientific understanding, prediction, and theory development.

*Speaker

Keywords: Machine Learning, Models, Representation, Toy Models, Understanding

Against Color Objectivism: A Critique from Electrodynamics

Xingyu Lyu * [1]

[1] University of Mannheim – Germany

This paper aims to argue against color objectivism by appealing to science, particularly electrodynamics. Color objectivism is the view that colors are properties of material objects. In the literature, the main versions of color objectivism can be roughly classified into two types: reflectance-grounded objectivism (including reflectance physicalism, microstructure physicalism and color primitivism) and color-disposition-grounded objectivism (including color dispositionalism, realizer and role functionalism). According to reflectance-grounded objectivism, colors are intrinsic, reflectance-related properties that are intrinsic properties of material objects, whereas reflectance properties are posited by science. According to color-disposition-grounded objectivism, colors are color-disposition-related properties such that the objects having those properties are disposed to look colored to certain perceivers in the relevant viewing conditions, where the color dispositions are realized by the underlying intrinsic properties of material objects, whatever science tells us those (intrinsic properties of material objects) are. Both types of color theories are claimed to be well compatible with science. However, an examination of electrodynamics does not suggest so. In this presentation, first, I will show two types of color objectivism share a common ground in how they incorporate the properties posited by science: The physical properties responsible for light reflection are intrinsic to material objects. Second, I will argue that color objectivism conceptually conflicts with electrodynamics, because the abovementioned common ground conflicts with the conceptual framework of electrodynamics. As a result, if electrodynamics is true, then color objectivism is false. Third, I will present a novel version of color physicalism that is compatible with electrodynamics under certain background assumptions.

Keywords: color objectivism, electrodynamics, light reflection

*Speaker

Can Complexity Science revolutionize our understanding of Deep Neural Networks?

Markus Maier *† [1]

[1] Hochschule für Philosophie München – Germany

My paper focuses on the question of whether Complexity Science can help us understand and extend our knowledge of Deep Neural Networks (DNNs). To do so, I will (1) discuss a recent proposal by Roberts, Yaida, and Hanin, who propose an "effective theory approach" to neural networks. I will place this idea in the larger historical context of complexity science and argue that it fits well within that tradition. (2) I will analyze this proposal from a philosophy of science perspective. This will include questions about what kind of understanding the authors hope to gain from their new approach and what exactly they claim to explain.
The lack of explainability of DNNs – the currently predominant artificial intelligence (AI) systems – is an extremely important issue with far-reaching theoretical and practical implications. This "black box" character of AI is a widely accepted narrative in academic and public discourse, and usually serves the purpose of intuitively conveying the idea that the system under consideration is opaque in some fundamental way. Scholars from various disciplines such as philosophy, computer science, and mathematics are trying to find ways to make the inner workings of DNNs more transparent and to "peek inside the box" – with limited success.

Complexity Science refers to a scientific enterprise that emerged long before the advent of DNNs to deal with systems that are opaque and seemingly too complex for us to understand. Although it is not a unified science, it has been highly successful in dealing with systems that cannot be described by more traditional frameworks. Common to all these approaches is the rejection of a paradigm of linearity and equilibrium in describing macroscopic real-world phenomena and systems. While these tools and methods are often developed for specific systems, a central claim of Complexity Science is that they can be applied to the description and study of other, seemingly disparate systems – independent of their specific material basis. Prime examples of this practice include insect colonies, the immune system, biological brains, and economic systems, among others.

Applying this logic to the opacity of DNNs, it seems legitimate to ask whether there are insights from Complexity Science that can help improve our understanding of these systems. Even though DNNs are not classical examples of complex systems, from a Complexity Science perspective there is no *a priori* reason to exclude artificial systems from its scope. As I will argue, the "effective theory approach" for understanding DNNs by Roberts, Yaida, and Hanin constitutes a prime example of the core idea of Complexity Science: The application of sophisticated tools, methods, and concepts to other systems that do not necessarily share material similarities.

*Speaker
†Corresponding author: markus.maier@hfph.de

Can Complexity Science revolutionize our understanding of DNNs? Or is it just a small contribution to a much larger problem? It is questions like these that are important as they promise to contribute to the larger issue of the explainability of AI on the one hand and the theoretical reflection on the practice of Complexity Science on the other.

Keywords: Complexity Science, Complex Systems, AI, Artificial Intelligence, Deep Neural Networks, Scientific Understanding, Scientific Explanation

Pregnancy as Agency

Anne Sophie Meincke * [1]

[1] University of Vienna – Austria

Pregnancy is widely regarded as passive, something that happens to women rather than being an active process they engage in. Feminist scholars have likened pregnancy to a disease or disability (Mullin 2005) and alleged that it deprives women of their agency altogether, making them "life's passive instrument" (de Beauvoir 1997). Common medical practices exacerbate this perceived passivity, further alienating women from their bodily experiences (Young 2005; Villarmea 2021b).
The issue of obstetric violence is gaining attention, with feminist philosophers exposing the misogynistic roots in Western metaphysics (e.g., Kingma 2020; Villarmea 2021a; Villarmea & Kelly 2020; Meincke 2022; Villarmea 2024). Women worldwide are reclaiming their autonomy in making medical decisions during pregnancy and childbirth. Despite these positive developments, the perception of pregnancy as a passive state continues to challenge the empowerment of pregnant women in prenatal care. Thus, it is essential to reconsider if pregnancy truly deprives women of agency.

My paper challenges the passivist view of pregnancy, proposing instead that pregnancy involves agency. The passivist view relies on questionable assumptions about both pregnancy and agency. By revising these assumptions in the light of insights from recent debates in the philosophy of biology and its intersections with analytic metaphysics, we can reframe pregnancy as an agentive process.

The paper begins by outlining the passivist view of pregnancy as described by de Beauvoir (1997). It then reviews contemporary feminist attempts to mitigate this view, such as emphasizing prenatal mothering practices (Nelson 1994) and acknowledging pregnant women's active bodily adjustments (Mullin 2005). Finding these attempts insufficient, I examine the concept of agency deployed by classic philosophy of action (Davidson 1963, 1971) but find it inadequate for pregnancy.

I then adopt an alternative approach: conceiving of agency as a biological capacity of organisms (Walsh 2015, Moreno & Mossio 2015; Di Paolo et al. 2017; Sultan et al. 2022). This concept of bio-agency lacks the rationalistic and masculine connotations of the traditional notion of agency and allows for multiple layers of agency. I complement this analysis by considering the metaphysics of pregnancy. The passivist image of pregnancy, I show, is facilitated by the still predominant ancient idea that a pregnant organism is a mere container for the foetus, which, however, is empirically untenable (Kingma 2019; Meincke 2022; Nuño de la Rosa et al. 2021).
I finally argue that combining the bio-agency approach with a Process View of pregnancy (Meincke 2020, 2022) supports a scientifically grounded view of pregnancy as an (inter-)active

*Speaker

process, enacted at various levels of a woman's existence, including the biological levels. Reinterpreting pregnancy as agency, in line with recent enactivist work in cognitive science (Quintero Martínez & De Jaegher 2020; Ciaunica et al. 2021), can counter alienating tendencies in pregnant women's self-perception and enhance their autonomy in prenatal care. It also underpins recent endeavours to uncover the neglected agentive nature of childbirth (Villarmea & Kelly 2020; Villarmea 2024).

Keywords: philosophy of biology, metaphysics of pregnancy, agency, organism, process ontology, feminist philosophy

Theoretical values between chemistry and linguistics: Ferdinand de Saussure's stoichiometry of sounds

Cornelis Menke * [1]

[1] Johannes Gutenberg - Universität Mainz = Johannes Gutenberg University – Germany

The investigation of the theoretical values informing theory choice has proven fruitful for the philosophical understanding of theory change in many disciplines; this applies both to contexts of theory acceptance and (especially more recently) to the pursuit of theories. In historical linguistics, however, the reconstruction of theoretical values is still a desideratum (Koerner 2006).

The paper focusses on a particularly interesting case of theory formation in 19th century historical linguistics: Ferdinand de Saussure's hypothetical reconstruction of the Indo-European vowel system. In his most important work published during his lifetime, the Mémoire (1879), de Saussure combined arecontruction of the vowel system with the postulation of three sounds that had disappeared in the course of language development. Although the work as a whole was considered outstanding, acceptance of the theory was largely limited to partial aspects. The theory only gained increasing acceptance over half a century later, when successors to the postulated sounds could be directly proven in Hittite.

The talk will focus on the reconstruction of the epistemic values that guided de Saussure's theorizing, as well as their reception. In particular, I shall argue for the three theses:

First, de Saussure's theory is characterized by theoretical values that are particularly well known from physics and chemistry (Kuhn 1977; Hoyningen-Huene 2013), and this in such a clear manifestation as one wouldn't expect in linguistics (and that surpasses other natural sciences). This applies in particular to the connectedness and unification of theoretical parts, and even the prediction novel phenomena. This finding goes far beyond the usual description of 19th century linguistics, whose methodological ideals are known to be oriented towards the natural sciences; however, this orientation towards the natural sciences is usually seen manifested in the doctrine of the exceptionlessness of sound laws, which the neogrammarian" research programm hailed.

Secondly, I would like to argue that this finding can possibly be understood as the result of a direct influence of chemical methods on linguistics. Several pieces of evidence point to this: The biographical background of de Saussure, whose father was a mineralogist and who had initially studied chemistry himself. Working on the Mémoir also coincided with the establishment of the periodic table (Mendeleev 1871); the empirical proof of one of the unknown elements postulated by the periodic table occurred in the same year as its publication (Scandium or ekaboron). Further, there are theoretical affinities and even terminological borrowings from

*Speaker

chemistry: de Saussure called the class of sounds postulated "coefficients sonantiques"; they have a function similar to the "coefficient stœchiométrique" (in the terminology of the time) in stoichiometric equations.

Thirdly, it will be argued that the long time that passed before the theory was accepted can also be explained with recourse to the epistemic values that characterize the theory. While the Mémoir's theoretical argumentation was considered exceptionally difficult to understand, direct empirical evidence was lacking for a long time - evidence of the kind that the neogrammarian program had called for (ironically, precisely in reference to the natural sciences).

Keywords: Values, theoretical values, linguistics, history of science, theory change

A Game-Theoretic Model of Self-Rationalizing Prejudice

Christoph Merdes * [1]

[1] Jagiellonian University – Poland

Stereotypes are a common heuristic device to navigate our social world. Unfortunately, they are also the source of many an injustice. One kind of such injustice is testimonial injustice, where an agent is unfairly discounted in their credibility when offering their testimony.
A common argument for the injustice of discounting someone's credibility based on a stereotype is the often crude and inaccurate character of our stereotypes. Many of them are, at best, wildly inaccurate generalizations, at worst plain nonsense and fiction. In such cases, it is easy to reject the application of a stereotype: It is unfair as it is irrational.
However, cases have been observed where a particular prejudicial stereotype has created its own reality, rendering itself an accurate generalization. In such cases, it is much less obvious why a good epistemic agent ought to refrain from applying it – worse, the accuracy suggests it might be irrational not to make use of it.
Some authors in the theory of epistemic injustice suggest that if such a stereotype is justified depends on its origin. We find this historical logic unsatisfactory for reasons we do not have the space here to discuss. Instead, we suggest a more structural understanding. This understanding is based on a simple game-theoretic model of testimony. One agent testifies, the other has to decide whether to trust or distrust.
The stereotype is a signal for coordination, and we suggest its injustice is constituted by signaling towards a Nash equilibrium which is asymmetric in payoffs. The model also enables us to analyze variations in payoff, including the inclusion of non-epistemic payoffs. By reference to the game-theoretic model, it is easy to disentangle a sense in which it is for the individual rational to follow the prejudice, while at the same time, the configuration as a whole remains inefficient, and qua our interpretation of the payoffs as epistemic, irrational.
We conjecture that this type of analysis might be useful not only for the theory of testimonial injustice, but also for scientists who sometimes face a dilemma when choosing research topics or reporting certain results. Empirically accurate stereotypes run the risk of becoming naturalzed, all the more when the cause is so subtle as in the case of self-realizing prejudice. A historical framing is sometimes problematic in this regard as well, as social types and stereotypes shift, sometimes rather quickly, and the origins and present application drift apart. A structural framing as ours instead lays bare the synchronic dynamic

*Speaker

and points out who benefits and who suffers, while also explaining the stability of an at least epistemically quite inefficient state.

Keywords: Testimonial Justice, Game Theory, Prejudice, Social Science

Trust networks: trusting science, trusting the media

Hanna Metzen * [1]

[1] Universität Bielefeld – Germany

Public trust in scientific expertise has become an important topic within philosophy of science. Contributions usually focus on the relation between lay people and scientists: For example, Irzik and Kurtulmus (2019) analyze the trust relation between "a member of the public" M and "an individual scientist or a collective body composed of scientists" S. They acknowledge that any information that is provided by S must be communicated, for example by the media. Misrepresenting information can hinder warranted trust in science. Yet, ultimately, the trust relata are M and S. Members of other institutions function as mediators, but they are not looked at as trustees or trustors themselves.
I will argue that while such analyses certainly help to understand public trust in scientific expertise, they are also incomplete. In many cases, networks of trust relations – *trust networks* – are crucial. In such networks, the trust one has in scientific experts depends on the trust in members of other institutions, like (science) journalists or political authorities. In my talk, I will focus on the media. First, I will argue that trust networks involving journalists are relevant to discussions on trust. Second, I will provide an account of trust in journalists and show how this affects trust in scientific experts.

In the first part, I will present scenarios in which trust networks need to be considered. One of these scenarios concerns the problem of judging expert's trustworthiness. While lay persons cannot directly assess scientific reasoning, they may be able to make second-order judgements, for example by assessing honesty (Anderson 2011). However, this has rightly been criticized as too optimistic regarding lay people's abilities (Barimah 2022). A solution is to involve non-experts who are in a better position to make assessments than ordinary lay people. Especially science journalists can take over such intermediary roles, resulting in more complex trust relations.

Another scenario is shared oversight. Especially when experts make public recommendations or provide policy-advice, public oversight is valuable. Yet, this cannot be too widespread or permanently, otherwise there is a risk of creating a climate of distrust. Oversight needs to be shared among trusted citizens in order to enable trust (Warren 1999). Again, science journalists can play this watchdog role, for example by identifying value judgements (Elliott 2019) or by pointing to potential biases. Both scenarios require some amount of trust in the media.
In the second part of my talk, I will therefore look at trust in science journalists. An account of their trustworthiness is based on features that also ground trust in science, like good will or integrity. However, there are important differences. Media reporting sometimes has a partisan nature that mirrors the segregation of political opinions, which is exacerbated by social media (Anderson 2011). Journalism scholars also point out that even adopting a more critical attitude

*Speaker

towards authorities from all sides can come with less trust in the media (Schudson 2022). For understanding public trust in scientific experts, philosophers need to pay more attention to what grounds trust in journalists.

Keywords: trust in science, science communication, science journalism, scientific expertise, trust networks

On Boolean Inferential Methods for the Establishment of Constitutive-Mechanistic Models in the Cognitive and Biological Sciences

Johannes Mierau * [1], Jens Harbecke *

[1], Sebastian Schmidt [1]

[1] Witten/Herdecke University – Germany

According to the "mechanistic approach" to the cognitive and biological sciences, scientific explanations succeed by analyzing the mechanisms that underlie a phenomenon or "constitute" it on several levels (cf. Bechtel and Richardson 1993; Glennan, 1996; Machamer, Darden and Craver 2000; Craver 2007). The explanatory norms that underlie mechanistic explanations have drawn much philosophical interest in the last three decades (Baumgartner and Gebharter 2016; Chirimuuta 2014; Craver 2007; Couch 2011; Gebharter 2017; Harbecke 2010; Huneman 2010).

In our presentation, we are concerned with the formal rules, methods, and algorithms capable of establishing mechanistic models that form the core of mechanistic explanations. Our goal is twofold: On the one hand, we offer an algorithm implemented in a functional Python script that transforms Boolean data tables obtained from tests on multi-level systems into causal-mechanistic models compatible with these tables. On the other hand, we defend several novel philosophical insights suggested by, and associated with, the solutions produced by this script. Both aims are closely related and depend on the respective other. With regard to the definition of mechanistic models, we follow the work of Couch (2011) and Harbecke (2010), who emphasize the link to the regularity approach of causation.

We draw our inspiration from the widely applied configurational comparative methods - Qualitative Comparative Analysis (QCA) and Coincidence Analysis (CNA) (cf. https://cran.r-project.org/package=cna and Dusa 2019). These methods offer strong and convincing solutions for Boolean causal inference problems. At the same time, these existing configurational comparative methods with their current parameter options cannot be adapted to produce adequate modeling solutions for data generated by causal-constitutive structures found in the real world.

On the basis of our alternative algorithm "mLCA" (multi-level coincidence analysis) (avaiable as a Python script on github https://github.com/user-jm/multi-lvl-Coincidence-Analysis) and its results for some prototypical examples of multi-level structures, we defend the following philosophical claims:

*Speaker

- In contrast to the existing QCA and CNA libraries, the novel mLCA script consistently generates adequate mechanistic models for multi-level causal-constitutive data tables, and it therefore demonstrates how the philosophical mechanistic approach can be put to work in the sciences.
- Inference methods generating causal-constitutive models require information about level assignments of the causal factors listed in the data tables produced by multi-level structures. Multi-level mechanistic models are not retrievable from coincidence data without such additional qualitative information.
- Due to certain adaptions of the methods for causal inference to the requirements of mechanistic modeling, and due to the ability to generate multi-level models (cf. claim 1), the number of solutions generated by mLCA grows radically as a function of the number of relevant causal factors listed in a data table, and the reduction of the set of solutions attainable with the information provided by the coincidence data tables often fails to produce a single solution.
- If further reductions are possible at all, they involve an ineliminable pragmatic ingredient, which has sharp consequences for the realistic ambitions of mechanistic explanatory projects.

Keywords: Mechanistic Explanation, Boolean Inference, Causal Inference, Models in Cognitive Science

Knowledge-Based Consensus – Evaluating Process or Outcomes? Why Not Both?

Boaz Miller * [1,2]

[1] Zefat academic college – Israel
[2] ACEPS, University of Johannesburg – South Africa

My talk addresses the question of how one can evaluate whether a scientific consensus is knowledge based or otherwise epistemically justified. I review two leading approaches in social epistemology of consensus and dissent, and argue that properly construed, they are complementary, rather than rival.
I start by reviewing three common approaches that will *not* do-"Do Your Own Research" (Ballantyne et al. 2022; Levy 2022), evaluating scientific method, and evaluating scientific credentials (Anderson 2011). I argue that they are impractical and assume a misguided model of science that overlooks the distributed and collaborative nature of research.

There are, however, two more promising approaches. One approach, advocated inter alia by Longino (2002) and Oreskes (2019), evaluates the social-epistemic processes of inquiry, deliberation, and closure that lead to the consensus formation; specifically, whether they are governed by critical social-epistemic norms. The second approach, advocated inter alia by Tucker (2001), Solomon (2001) and Miller (2013; 2019; 2021; 2024), evaluates the consensus itself and its object; specifically, whether it exhibits apparent consilience of evidence and relevant social diversity.

While the process-based approach and the outcome-based approach are sometimes seen as rivals, I argue that they are better construed as complementary for two reasons. First, they are useful in different stages of inquiry. The process-based approach evaluates the stages of inquiry that lead up to a consensus. It may help researchers improve their inquiry and deliberation before a consensus is reached. By contrast, the outcome-based approach focuses on a later stage, in which a consensus has already been reached and imagines a decision maker who needs to decide whether to rely on it. Second, in practice, evaluating outcomes and evaluating process often boil down to examining the same substantive factors, such as the social and scientific backgrounds of the parties to the consensus and the range of evidence considered by them.
Nevertheless, I note two caveats regarding the process-based approach. First, when reviewing process, we should not assumed that science is a consensus-aiming enterprise. The same factors that generate a consensus may also generate a dissent (Solomon 2001). When a stable, long-lasting consensus is reached (Vickers 2022), social agreement is a byproduct of the stabilization of evidence and the congelation of facts, rather than science's fulfilling its telos. Thus, when facing a consensus, we should still ask whether knowledge, as opposed to non-epistemic influences, is its best explanation. Second, the notion of consensus should not be built into the very definition of knowledge. That is, knowledge should not be defined as an agreement reached by following a good process. A good process doesn't guarantee a good outcome: a community

*Speaker

that follows best social-epistemic practices may still err or fail to rule out a relevant alternative theory, hence fail to reach knowledge. Bearing these two caveats in mind, the process-based approach and the outcome-based approach can be fruitfully combined for assessing a scientific consensus for knowledge.

Keywords: consensus, dissent, knowledge, social epistemology

Epistemic Rehabilitation. First Steps

Daniel Minkin *† [1]

[1] Institut für Philosophie, Justus-Liebig-Universität Gießen – Germany

In the philosophy of science, there are numerous studies on what constitutes scientific misconduct. In contrast, relatively little work has been done concerning the question of what should be done *after* the misconduct has happened. Should we exclude those responsible from the community of epistemically trustworthy agents once and for all? Or should we adopt the attitude that everyone deserves a second chance-regardless of the severity of the misconduct? Neither option seems feasible, since the question of whether an "epistemic rehabilitation" is appropriate depends on many factors. In my paper, I want to examine these factors more closely.
In the first part, I will present some general thoughts on epistemic rehabilitation. I begin with a definition of an initial condition, which goes as follows:

(IC) An epistemic agent X became epistemically untrustworthy at time t due to X's acts.

On the basis of (IC), I formulate the following definition of "epistemic rehabilitation":

(ER) X is epistemically rehabilitated iff X becomes epistemically trustworthy again at t+1.

By "epistemic trustworthiness" I mean the presence and by "epistemic untrustworthiness" the absence of sufficient reasons to epistemically trust X's testimony. By "epistemic trust" I mean the readiness to take X's testimony as a basis for improving one's own web of belief.

Against this background, I will formulate my central question:

(CQ) Under what conditions are we entitled to treat X as epistemically rehabilitated?

With regard to (CQ), I want to put forward three intertwined assumptions:

a) (CQ) can only be answered by a case-by-case study.

b) No set of sufficient criteria for treating X as epistemically rehabilitated applies to every case.

c) A well-devised epistemological rehabilitation program that fits all cases is not possible.

In the second part, I will analyze three examples of scientific misconduct to argue for these three assumptions:

*Speaker
†Corresponding author: danminkin@msn.com

First case: In 1998, Andrew Wakefield and twelve other authors published a study suggesting a link between the MMR vaccine and autism. In 2006, the British newspaper *Sunday Times* revealed that Wakefield and five other authors had received more than 500000 $ for providing evidence in favour of the link. Although later studies could not corroborate Wakefield's assumption, he continued to defend them.

Second case: In 2011, the University of Tilburg proved that Diederik Stapel, a former dean of the university, had manipulated and fabricated data in more than 50 socio-psychological studies on various topics. After his dismissal, Stapel admitted to the fraud and promised that he would contribute to uncovering scientific misconduct in the future.
Third case: In 2019, the University of Tübingen and the German Research Foundation launched an investigation into a study by Niels Birbaumer. Birbaumer and his team worked on a Brain-Computer-Interface that enabled amyotrophic lateral sclerosis patients in a completely locked-in state to communicate. The investigation found that the researchers withheld data, thereby distorting the study's conclusion. A later study, in which Birbaumer was also involved, provided support for the earlier findings.

Keywords: Scientific Misconduct, Fraud, Trust, Science and Values, Epistemic Rehabilitation

Between Autonomy and Entanglement: Middle Path Frameworks for the Organism-Environment Relationship

Katie Morrow *† [1], Marie I. Kaiser [1]

[1] Universität Bielefeld – Germany

This paper critically evaluates three strong views about the organism-environment relationship in biology: one on which environments are autonomous, one on which organisms are autonomous, and one on which organisms and environments are entangled. While these schools of thought have been theoretically influential, they face clear shortcomings in application to empirical studies on behavior and ecology. In view of this, we critically survey a series of proposed "middle path frameworks" that attempt to recognize the mutual causal influences of organisms and environments while still maintaining that individual organisms can be identified and delimited within biological research.

The concept of *autonomy* has informed several influential schools of thought in the philosophy and theory of biology. Traditional views of natural selection conceive the environment as an autonomous force that acts like a sieve to filter organisms. On this *externalist* view of natural selection, the environment is a pre-existing, external force to which the organism adapts (discussion in Lewontin 1982; Levins and Lewontin 1985; Lewontin and Levins 2007).

Separately, there has been a surge of interest in autonomy as a concept for conceptualizing organisms or living systems in the philosophy of biology (e.g., Moreno and Mossio 2015; Arnellos 2016; Bich and Bechtel 2021). Biologists and philosophers with a focus on the autonomy of organisms have emphasized the ways in which organisms are distinct from complex abiotic processes, maintain their identify over time, and exhibit agency.

These perspectives on autonomy contrast with work emphasizing the mutual influences and interdependence of organisms and environments (von Uexküll 1909; Haldane 1931; 1935; Sterelny 2005; Aaby 2021; Walsh 2022). For instance, proponents of niche construction theory emphasize that while environments act on organisms via natural selection, organisms reciprocally alter their environments via niche construction (Odling-Smee, Laland, and Feldman 1996; Laland, Odling-Smee, and Feldman 2019). This perspective has led some theorists to conclude that organisms and environments are strongly causally or ontologically *entangled*, even to the extent that they cannot be separated.

These views are in tension with practice in e.g. behavioral biology and organismal ecology, in which it is presupposed that organisms and environments exert mutual influence, yet individual organisms must be delimitable for research purposes. We critically survey a variety of frameworks that attempt to walk a middle path between the extremes of autonomy and entanglement.

*Speaker

†Corresponding author: katherine.morrow@uni-bielefeld.de

The frameworks we consider include proposals relating to reciprocal causality (Baedke 2019), the notion of environmental affordances (Walsh 2015), the extended phenotype (Edelaar, Otsuka, and Luque 2023), individualized niches (Takola and Schielzeth 2022), mechanistic explanation, and process ontology. We evaluate the proposals especially with an eye to their applicability in the context of behavioral and ecological research. Some of the frameworks are useful in particular research contexts, while others are of philosophical interest while being less applicable to empirical research. In general, there is a lack of a middle path framework that is widely applicable across biological fields, suggesting an opportunity for further philosophical/theoretical work.

Keywords: biology, organism, autonomy, entanglement, niche construction, extended phenotype, reciprocal causality, individuality, behavior, ecology

Are Plants Cognitive Agents? A Naturalistic Approach to Comparative Cognition and Plant Behavior Studies

Filippo Murabito *† [1], Emma Braccini * ‡ [2]

[1] Università degli Studi di Firenze = University of Florence = Université de Florence – Italy
[2] Technische Universität Dresden = Dresden University of Technology – Germany

The characterization of cognition remains problematic: despite significant efforts, no definitive criteria have been established for identifying this concept. As a result, there is no unanimous agreement on which life forms should be classified as cognitive and which should not. While humans are typically regarded as the paradigmatic model of cognition, growing evidence suggests that cognitive processes extend far beyond the animal kingdom, potentially encompassing all forms of life. Examples include decision-making at the cellular level, social behaviors in bacteria, and learning abilities in slime molds. The emerging field of Basal Cognition seeks to develop a unified framework to investigate cognition across a broad range of organisms, from those without nervous systems to those with neurons, neural networks, and simpler centralized nervous structures. This approach aims to examine cognition through an evolutionary lens, viewing it as a systemic biological function crucial for the survival, well-being, and reproduction of any autonomous living system. The central hypothesis is that cognitive functions exist on a biological continuum shared by all organisms. My paper, which aligns with this research field, seeks to address the following question: Do plants exhibit the behavioral traits required to attribute cognitive agency? The majority of arguments against the idea that plants could be cognitive agents follow a similar structure: 1) Cognitive organisms exhibit flexible behavior; 2) Plant behavior is inflexible and hard-wired; 3) Therefore, plants are not cognitive. I aim to critically assess these arguments from the standpoint of a philosophy based on empirical evidence. While I agree that rigid, hard-wired responses to environmental stimuli are not particularly compelling from a cognitive perspective, recent empirical findings indicate that plant behavior is far more complex than mere reflexive actions. Plants exhibit adaptive, flexible, anticipatory, and goal-directed behaviors. Indeed, plants are capable of decision-making and risk assessment, evaluating various environmental factors, and selecting responses that best address their immediate needs. Moreover, although plants are often viewed as stationary, they actually display a wide range of movements, many of which are goal-directed and driven by intentionality. Plants also possess memory and the ability to learn. Memory is vital for plants, enabling them to survive by recalling and integrating past experiences with present conditions to predict future outcomes. Learning, which can range from simple to more complex processes, is fundamental to memory formation. Based on this evidence, I propose that plant behavior shares significant similarities with animal cognitive behavior, suggesting that plants could be regarded as cognitive agents in a literal, non-metaphorical sense. Thus, plants possess the traits required to be considered

*Speaker
†Corresponding author: murabitofilippo17111997@gmail.com
‡Corresponding author: emma.braccini@gmail.com

potential cognitive agents. This makes them valuable as experimental models for comparative studies of cognition. By adopting non-animal models-such as plant models-we can advance the discussion about the nature of cognition and may discover that brains are not necessarily required to perform, at least, certain cognitive behaviors. However, much more research is needed to better understand plant behavior and the mechanisms driving it.

Keywords: cognitive sciences, comparative cognition, plant cognition, basal cognition, philosophy of cognitive sciences

Capacity building for responsible climate research

Francesco Nappo * [1]

[1] Politecnico di Milano – Italy

Scientific laboratories and institutions conducting or assessing economic research on climate change are under increasing pressure to recognize the societal responsibilities associated with their work. This includes, among other things, being transparent about the values that underlie economic modeling choices and taking responsibility for the consequences that those influences might have (Douglas 2009). However, the complex nature of contemporary climate science presents important challenges. Large workflows, strict deadlines, diverse data sources, and the growing number and complexity of climate models contribute to making the oversight and management of value-laden scientific practices an open problem for most actors involved.
To ensure a proper response by the relevant scientific community, we put forward *capacity building* as a priority. Research infrastructures with specific strengths to oversee complex workflows and hold actors accountable are needed to preserve the reliability of public and private climate economic research. The philosophical task consists in outlining the type of capacity building that is needed. In this regard, we first discuss a recent argument by Douglas (2023), which suggests that there is an important risk of politicizing science if large scientific institutions are put in charge of overseeing responsibilities for broad societal impact. In agreement with Douglas, we argue that a soft governance mechanisms by *mid-range* infrastructures, such as dedicated laboratories or dedicated divisions of national research centers on climate change, rather than large institutions such as the IAMC, would best fulfill current capacity needs.

Secondly, with regards to the specific tasks to be undertaken, we distinguish two kinds. *Adversarial* mechanisms adopt established or emerging methods to provide information about databases and models with the aim of checking potential divergences from co-designed goals. An example of an adversarial tool is the application of Monte Carlo simulations and machine learning analysis techniques to emission scenario ensembles (Bosetti et al. 2016). Conversely, *assistive* mechanisms possess a broadly generative function, in providing researchers with new methods and visions to address emerging stakeholders' and end-users' interests. For example, developing new tools for operationalizing and addressing issues of justice in scenarios or new model intercomparison methods is a highly desirable goal (Zimm et al. 2024). Based on this distinction, we put forward concrete suggestions as to the kinds of incentives and funding schemes that can help realize the required capacities within current institutional arrangements.

Bosetti, V., Marangoni, G., Borgonovo, E., Diaz Anadon, L., Barron, R., McJeon, H., Politis, S., Friley, P. 2015. Sensitivity to energy technology costs: A multi-model comparison analysis. *Energy Policy.* 80(May 2015): 244-263.

*Speaker

Douglas, H. 2009. *Science, Policy, and the Value-Free Ideal.* Pittsburgh: University of Pittsburgh Press.

Douglas, H. 2023. Structuring institutions for responsible and accountable science. *Philosophy of Science.* doi:10.1017/psa.2023.128
Zimm, C., Mintz-Woo, K., Brutschin, E., Hanger-Kopp, S., Hoffmann, R., Kikstra, J., Kuhn, M., Min, J., Muttarak, R., Pachauri, S., Patange, O., Riahi, K., and Shinko, T., 2024. Justice considerations in climate change research. *Nature Climate Change.* 14: 22-30.

Keywords: climate change research, value management, climate economics, scientific institutions, research ethics

The role of phenomenal consciousness in knowledge – what we did not talk about

Saskia Janina Neumann * [1]

[1] Eötvös Loránd University – Hungary

Suppose you wake up in the hospital feeling normal. The doctor holds her hand in front of your face and asks you to count how many fingers she is holding up. You answer correctly and she gives you some candy as a reward. Now she holds her hand just outside your conscious field of vision and asks you to count how many fingers she is holding up. You tell her you can't see her hand, but she asks you to guess anyway. You say, "What the hell" and hazard a guess. To your surprise, she says you got the right answer and gives you some more candy. Again she holds up her hand and asks you to guess the number of fingers, and again you give the right answer. The experiment continues until you've eaten so much candy you begin to feel sick (Smithies, 2019, 77f.)

A situation like this could happen to a person with an impairment called blindsight. Declan Smithies (2019) argues for the view that people with blindsight do not have knowledge in these situations as they do not have any phenomenal access to their guesses. I will argue for the view that we are all like blindsighted people as we all lack phenomenal access to our guesses and even, worse, to all of our beliefs!

Yet, subsequently, I will also claim that all is not lost. Whilst we do not have phenomenal access to our beliefs, we can still have knowledge and even phenomenal access can be useful but only at the point in time when we reflect on our beliefs, not at a later time. In my talk, I will firstly introduce Smithies view and will secondly argue against it, based on empirical evidence. Thirdly, I will argue for the view that we still do have knowledge and fourthly explain how phenomenal access to our beliefs is useful after all, at least, when it comes to maintaining knowledge.

Keywords: Phenomenal Consciousness, Consciousness, Philosophy of Mind, Philosophy of Cognitive Science, Knowledge, Epistemology, Philosophy of Epistemology

*Speaker

Is the Universe Fundamentally a Density Matrix?

Alyssa Ney * [1]

[1] LMU Münich – Germany

This paper examines the case for density matrix realism over wave function realism, as an approach to the fundamental ontology of our world. To date, there are two arguments that have been used to motivate density matrix realism. One is that we get a simpler metaphysics if we move from wave function realism to density matrix realism. The second is that density matrices are more general than wave functions, in allowing that the universe could be in a mixed state. To be convincing, these arguments can't rest on the logical possibility that our universe could be in a mixed state. After all, if we assume standard understandings of mixedness, it is a contradiction to say that the universe as a whole could (objectively) be in a mixed state. Thus, one would need to appeal to some scientifically-backed argument to revise the conception of mixedness. Here, Hawking's model of black hole evaporation seems poised to come to the rescue, as it implies the universe will evolve into a state that is mixed, and so we develop an argument to this effect. Although we find the prospect of using black hole evaporation to argue for density matrix realism to be intriguing and worthy of investigation, we show there is substantial reason to be cautious. Most physicists, including Hawking, believe that this model will be superseded by a more fundamental theory of quantum gravity. Despite the lack of consensus on the details of such a theory, the guiding principles that do enjoy some consensus point away from an interpretation of black hole evaporation in which the universe as a whole evolves into an (objectively) mixed state. We conclude that there do not seem to be compelling motivations coming from the most sophisticated candidates for fundamental physics to entertain situations in which the universe is characterized by a density matrix rather than by a wave function or state vector.

Keywords: quantum ontology, wave function realism, density matrix realism, open systems, black hole evaporation

*Speaker

To Spin or Not To Spin: A Dogma of Bohmianism Revisited

Andrea Oldofredi *† 1

1 University of Lisbon – Portugal

As it is very well-known, Bohmian Mechanics (BM) is a deterministic quantum theory of particles which move in three-dimensional physical space and follow continuous trajectories. This theoretical framework is statistically equivalent to the standard formulation of quantum mechanics although their physical content is remarkably different, since the former makes a clear metaphysical hypothesis concerning the intrinsic corpuscular nature of matter.
Referring to this, according to the formulation of Bohmian Mechanics (BM) due to Dürr, Goldstein and Zanghì the *only* property instantiated by quantum particles is their position, i.e. their exact location in a background space. From position some derivative quantities can be defined, as for instance velocity and momentum. On this perspective, however, other quantum observables are not considered real or genuine attributes of the Bohmian corpuscles. In particular, it has been argued in several places that spin does not exist in this theory, or better that it does not refer to any physical property of the Bohmian particles. Moreover, since one of the main philosophical tenets of the primitive ontology program - the general framework in which BM is usually presented - is to be realist only towards those entities playing a fundamental explanatory role, as Allori claimed, and given that spin measurements are said to be reducible to position measurements in BM, it is concluded that spin must not be real.

The non-existence of spin is an undisputed belief held by the overwhelming majority of Bohmians, to the point that it has become a *dogma of Bohmianism.*

In this talk I critically assess the theoretical basis behind this "dogma". Contrary to the received view, I provide arguments for the reality of spin in Bohmian mechanics based on recent results in Bohmian quantum chemistry, where spin-dependent trajectories are routinely employed. In particular, I show that assuming the existence of spin one has significant explanatory advantages over canonical BM. In fact, among other things, the standard guiding equation of BM entails that in the Hydrogen ground state the velocity of the electron around the nucleus is zero; therefore, the electron is at rest in such a state. On the contrary, taking into account the momentum of a Bohmian particle including spin components, such a counterintuitive feature of Bohmian mechanics vanishes, for it is shown that the electron *does orbit* around the nuclei. It will be argued in addition that this feature significantly improves the explanation of the chemical bond w.r.t. standard BM.

If employing spin-dependent guiding laws in BM entails major explanatory advantages, and if one must be committed to the reality of those theoretical entities playing a crucial explanatory role, then one should admit the existence of spin as a genuine property of the Bohmian

*Speaker
†Corresponding author: aoldofredi@letras.ulisboa.pt

particles.
To this end, following Holland, Vigier and Dewdney, I will argue that although spin measurements can be described in terms of position measurements, such reduction is by no means eliminativist, contrary to the usual claims made by contemporary Bohmians, and thereby against their dogma.

Keywords: Bohmian Mechanics, Spin, Trajectories, Chemical Bond

Performative Paternalism

Jakob Ortmann * [1,2]

[1] Department of History and Philosophy of Science, University of Cambridge – United Kingdom
[2] Leibniz Universität Hannover=Leibniz University Hannover – Germany

It is a well-appreciated fact in sociology and philosophy of science that scientific conceptualisations, broadly construed, do not merely *depict* their objects of study, but can sometimes *change* them. For example, if a prominent economist publicly announces that they think a bank run is likely, this announcement may very well magnify the probability of the bank run actually occurring, making it a "self-fulfilling prophecy". A common thread among the broader literature is that science's causal influence on the very targets it is studying tends to create *epistemic* and *ethical* issues. Epistemic issues, because it is not clear whether in such cases there even exists "a stable object (...) to have knowledge about" (Hacking 1995, 61). Ethical issues, because it is not self-evident what the scientist should do with such unsought transformative power; for example, should the economist have kept quiet about a possible bank run or did they have a responsibility to warn the public?

Various strategies have been proposed to handle both problems, the ethical and the epistemic. According to what I will call here the *strategy of manipulation*, we need not be too troubled by epistemic inadequacies caused by performativity. Instead, the argument goes, we could focus on the consequences of performativity from an ethical point of view and use performativity as a tool for eliciting changes or states within the target that are deemed desirable. Variants of this strategy have been applied to epidemiology (van Basshuysen et al. 2021), climate change negotiations (Ortmann and Veit 2023), decolonial activist research (Koskinen 2022), human kinds more generally (Godman and Marchionni 2022) and more.

The strategy of manipulation has provoked various criticisms questioning its legitimacy. Specifically, it gave rise to a concern I will dub *performative paternalism.* Winsberg and Harvard (2022, 516), for example, argued that using performativity as a tool to drive certain outcomes would be unjust and deceitful. Concealed as a value-neutral stance of just "following the science", deliberate deployment of performative models would impose scientists' own value judgments on society. By adopting a manipulative role, scientists act outside their legitimate competencies. As such, they argue, performativity "is never a legitimate purpose for a model", and "a serious threat to democratic decision making". Khosrowi (2023) and van Basshuysen et al. (2021, 122-123) share a similar worry.

Contra these worries, in this essay, I contend that the ethics of the manipulation strategy are more convoluted. As such, there are conceivable conditions under which the strategy of manipulation can be justified. Specifically, I argue for two things. (i) Autonomy-infringing deceit is not a necessary feature of the strategy of manipulation and, therefore, it does not have to be paternalistic. (ii) In cases where a performative model choice is inescapable, making it paternalistically can be justified. As such, this argument contributes to a growing literature on performativity, which has made notable advances regarding its epistemic aspects but has left

*Speaker

ethical aspects underdeveloped, yet (Marchionni, Zahle, and Godman 2024, 8).

Keywords: Performativity, reflexivity, paternalism, values

A History of the System

Tuomas Pernu * [1]

[1] University of Eastern Finland – Finland

The notion of system is essential to all of science. But what are systems? Is the term "system" referring to the same thing in all the wide range of disciplines it is being used in? There is no philosophical discussion addressing these questions. Recently, however, they have received attention in engineering (Hitchins 2009; Dori & Sillitto 2017; Sillitto *& al.* 2017; Yang *& al.* 2019; Dori *& al.* 2020; Kasianiuk 2021; Salado & Kulkarni 2021). This is not surprising, given that the aims of engineering are in providing us with efficient and robust devices – systems – for the manipulation and control of nature.

Here, the issue of defining "system" is approached historically, by asking: through what kind of developments have we arrived at the current variety of usages of the notion? Five stages are identified. First, the etymological roots of the term are in antiquity, and the origins of the idea of control or feedback system can be traced to these times (Mayr 1970). Second, important elements of the modern notion were defined during the enlightenment, by Cartesian geometry and Newtonian mechanics. Third, a decisive step was taken during the 1900th century, with the rise of thermodynamics and statistical mechanics; the term "system" becomes ubiquitously used in mathematics and physics. Fourth, systems come in explicit focus in the mid-20th century, with the rise of cybernetics (Wiener 1948) and systems theory (von Bertalanffy 1968). Finally, in the 21st century a variety of methods, in a wide range of fields, are aimed at analysing complex systems (Ladyman & Wiesner 2020).

Can this evolution be viewed as a continuous and consistent narrative? Not without heavy reinterpretation. The system theoretical (entity based) notions, currently in focus in engineering and complexity analyses, seem fundamentally different to the abstract (phase space based) notions used in mathematics and physics. Perhaps unifying these views is possible. Perhaps – but not without acknowledging the historical roots of the differences.

Keywords: control theory, cybernetics, feedback, complex systems, systems theory

*Speaker

The transfer of game-theoretic models: formal templates, unification, and scientific progress

Edoardo Peruzzi * [1]

[1] Leibniz University Hannover – Germany

Model transfer, the process of applying models or modeling techniques across different scientific domains, has recently been conceptualized by philosophers of science through the notion of a template. Templates are formal and mathematical structures that serve as the basis for constructing models that can be applied across various fields. Paul Humphreys' template-based approach is particularly insightful because it addresses three key questions about model transfer: (i) what is the *unit of transfer*, (ii) what makes it *transferable*, and (iii) what *empirically justifies* the transfer.

Humphreys argues that the unit of transfer is a *formal template*-a general, mathematically based structure that can underpin model construction in multiple scientific domains. Formal templates are transferable due to two main factors: the construction process itself and the mathematical or computational tractability of the formal template. Formal templates are not specific to any scientific field, as they often rely on assumptions rooted purely in mathematics, such as the exponential growth equation, or have been abstracted from their original theoretical context, such as the Ising model, or the Lotka-Volterra equations. Finally, the empirical justification for template transfer depends on the satisfaction of the construction assumptions. For example, in exponential growth models, the assumption of a constant growth rate might not hold for human populations, where factors like resource limitations and social behaviors come into play. However, the same model can accurately describe bacterial reproduction in controlled laboratory conditions, where such assumptions are satisfied.

One of the most prominent examples of model transfer in the social and behavioral sciences is the widespread diffusion of game-theoretic models in economics, political science, biology, and beyond. It is natural, then, to ask whether this case of model transfer can be understood through the template-based account developed by Humphreys. Although Humphreys' analysis primarily focuses on formal templates like differential equations, I contend that it is indeed possible to apply these concepts to models in different formats, such as game-theoretic models. Thus, my first contribution is to demonstrate how game-theoretic models function as formal templates within Humphreys' framework.

Despite its merits, however, Humphreys' template-based account has a significant limitation when applied to the transfer of game-theoretic models. His approach offers a *static* view, emphasizing the initial construction of templates for transfer without addressing how these templates evolve when applied in new disciplinary contexts. Although he acknowledges that templates

*Speaker

are often adapted or refined during transfer, he does not explore this process in detail. My second contribution, therefore, is to refine Humphreys' framework by introducing the concept of *cumulative model transfer*. In this process, formal templates are not only adopted within a particular domain but also modified to address domain-specific challenges. These modified templates can then be transferred to other fields, where they are further adapted for constructing new models. Using examples such as Bayesian games in economics and political science and the Replicator Dynamics in evolutionary game theory, I illustrate how cumulative model transfer operates. These examples demonstrate how model transfer contributes to scientific progress by fostering explanatory and disciplinary unification.

Keywords: model transfer, game theory, formal template, scientific progress, unification

Exclusion, interventionism, and theoretical interests

Niccolò Aimone Pisano * [1]

[1] University of Stirling – United Kingdom

The core idea of the causal exclusion problem is that, if one is committed to a rather minimal form of nonreductive physicalism, and if some plausible (although not uncontroversial) principles are granted, then one has to admit that higher-level properties in general, and mental properties specifically, are causally inert.
Recently, there have been several attempts at offering solutions to this problem based on the adoption of James Woodward's (2003) interventionist account of causation. However, it has also been argued (Baumgartner (2018)) that the exclusion problem cannot be dispelled as a result of the adoption of interventionism: this would only be the case if the supervenience bases of the relevant higher-level variables were independently fixable, which they are not for metaphysical reasons.

The focus of my talk is on Woodward's own solution (2015, 2022). This particular proposal revolves around an amended version of Woodward's original account of interventionism. The main difference between this "interventionism*" and the original version is that the latter is only concerned with purely causal graphs, while the former explicitly takes into account causal graphs also involving metaphysical determination relations different from causation. Specifically, interventionism* highlights the need *not* to hold the supervenience bases fixed while performing interventions on the relevant higher-level supervening variables, thus avoiding the aforementioned "independent fixability objection".
The main claim that I advance is that the reasons supporting interventionism* cannot be accepted in the context of the debate over the exclusion problem. This is not to say that one should not adopt interventionism* for pragmatic purposes in ordinary causal reasoning. Rather, when it comes to the exclusion problem, the considerations playing a role in deciding whether some variable is to be held fixed or not tilt the scale in favour of the need to hold the variables representing supervenience bases fixed, when an intervention is performed on the variables representing the relevant supervening (mental) properties. Failing to do so would make it impossible to answer the causal questions that motivate the discussion around the exclusion problem itself in metaphysics, specifically: "Do higher-level (mental) entities possess causal powers distinct from those of their supervenience bases?" (notwithstanding Jessica Wilson's, 2005, 2011 view that the right question is rather "Do higher-level (mental) properties possess causal *profiles* distinct from those of their supervenience bases?"). But, as Woodward (2022, p.22) maintains, the causal questions one is interested in answering are part of the contextual factors guiding the decision to consider some variable as a confounder or not, and hence the choice of whether to hold it fixed. Therefore, failing to hold the supervenience bases fixed, as recommended by interventionism*, prevents one from ascertaining whether the causal powers that one may attribute

*Speaker

to mental properties are also thereby possessed by the relevant supervenience bases, or whether they are indeed distinct. This is unacceptable in a metaphysical context where causal powers are taken to motivate existential claims (because of principles such as Alexander's Dictum, according to which a property is real iff it possesses causal powers).

Keywords: Causal exclusion, Interventionism, Nonreductionism, James Woodward

Democratic Deliberation and Values in Science: Lessons from Systemic Approaches

Elena Popa *† [1], Michał Zabdyr-Jamróz [2]

[1] Jagiellonian University – Poland
[2] Jagiellonian University Medical College – Poland

Recent work on the role of values in science has pointed to deliberation as a way of reaching politically legitimate decisions. For instance, seeking alignment with the values of the public has been suggested as a solution for problems involving the distribution of inductive risks in scientific decision-making (Irzik & Kurtulmus 2019). There have also been suggestions of subjecting value choices by scientists to public deliberation (Schroeder 2021). While engaging with the values of the public through mini-publics or citizens' assembies is an appealing suggestion, this proposal has also attracted criticism. Work on the rationality of distrusting experts alongside more specific discussions of vaccine hesitancy have highlighted that there are problems with engaging groups that have persistently been excluded (Dujif 2021; Ivani & Dutilh Novaes 2022). Deliberation has also been criticized for focusing too much on the ideal of reaching consensus, which can exclude particular groups or individuals (Van Bouwel & Van Oudheusden 2017; Koskinen 2022). Additional challenges stem from so-called 'partisan science' projects which pursue specific values (e.g., feminist) which would not necessarily be likely to bee agreed upon through a deliberation process (Hilligardt 2024), as well as cases for pluralism as opposed to democratic alignment in the case of social scences (Thoma 2024). This paper will draw on work from political philosophy, particularly systemic approaches from deliberation and democratic theory, to specify the role of deliberation in connection to value choices. Systemic approaches (e.g., Mansbridge et al. 2012) shift the focus from individual deliberation sites (such as one mini-public) to their broader context, involving wider social dynamics. This enables a more in-depth understanding of particular problems arising in deliberation and solutions that engage the relevant social dynamics and actors. A key point is that on a systemic approach the role of deliberation is not always to reach consensus, and deliberations can serve an educational purpose or help empower participants. This helps deal with the objections to the consensus ideal. Attention to the broader context and conflicting interests can also help deal with issues such as hijacking - pointing to structural features that impede a fair deliberation process (e.g., the influence of industry funding in science). Lastly, even in cases where consensus is impossible and one settles for agonistic pluralism, deliberation can help participants gain more awareness of their own position and those of others. This can help in difficult cases where there is no one right solution cutting across divergences in values and interests: once a choice is made in accordance to one set of values, future needs in connection to the values set aside are highlighted to be prioritized in future choices.

*Speaker
†Corresponding author: elena.popa@protonmail.com

Keywords: values in science, deliberation, deliberative stance, science and democracy, pluralism

Idealization and Abstraction in Scientific Modeling

Demetris Portides *† [1], Athanasios Raftopoulos [2]

[1] University of Cyprus [Nicosia] – Cyprus
[2] University of Cyprus – Cyprus

It is commonplace that Science doesn't work by describing the target systems of the world in their full complexity, but instead some general principles that are believed to be non-accidental generalizations are used as antecedents that guide the construction of models of physical systems. Such models are thought to be among the primary instruments of scientific representation of target physical systems.
In scientific modeling simplifying assumptions enter in a variety of different ways, e.g. by ignoring physical parameters, or by considering the dimensions of a body to be infinitesimally small, or by considering a variable quantity to have continuous values, or by considering the quantity of a parameter to be uniformly distributed etc. The class of model-simplifying assumptions seem to include heterogeneous members.

One possible attempt to bring order to the heterogeneity of this spectrum is the frequent admission that model-simplifying assumptions involve two rather distinct characteristics: the omission of features of the phenomena from the scientific representation of respective target systems (often referred to as *abstraction*) and the modification of features retained in the representation (often referred to as *idealization*).

There seems to be a difference between these two kinds of assumptions, even though both of them lead to a descriptive caricature of the world. This apparent difference dictates a question: how can the two kinds of assumptions be defined so that the given definitions lead to a clear-cut distinction of the two?

Some philosophers pay particular attention to the question and attempt to spell-out particular ways by which to explicate the distinction. Such examples include Cartwright (1989), Jones (2005), Godfrey-Smith (2009), Levy and Bechtel (2013). In this work, we will defend the thesis that although it is possible to distinguish idealization from abstraction as two facets of the same cognitive act, the attempts by the latter group of authors to distinguish the two fail to meet the desired goal.
We explain some of the conceptions of the distinction between abstraction and idealization. Being entirely aware of certain differences among the various conceptions we ignore them in order to focus on their main common characteristics, as this work is not meant to be a detailed review of the relevant literature. We put forward some arguments why these conceptions fail to meet their purpose, which is to offer a clear way by which to distinguish abstraction and

*Speaker
†Corresponding author: portides@ucy.ac.cy

idealization. Finally, we suggest a different way by which to distinguish the two: that both idealization and abstraction could be understood as particular forms of the same cognitive act (or thought process)-that of *selective attention*. By attending selectively to particular aspects of the target system modelers abstract away from unwanted noise, and by attending selectively to particular features of those aspects, or in particular ways, modelers idealize. The same cognitive act employed in different ways is responsible for the two ostensibly distinct categories of model-simplifying assumptions.

Keywords: models, idealization, abstraction

Practical and theoretical understanding through exploratory design

Michael Poznic * [1], Christopher Pincock [2]

[1] Karlsruhe Institute of Technology = Karlsruher Institut für Technologie – Germany
[2] The Ohio State University [Columbus] – United States

This paper draws on scholarship on exploratory experiments (Steinle 2005, 2016) and exploratory models (Gelfert 2016, 2019) to consider philosophical issues raised by exploratory design. Exploratory design focuses on how to best make a new sort of artifact such as an improved airplane wing. When this investigation is not primarily guided by settled theory, the design process is exploratory in the sense that it involves mapping out a range of possibilities and selecting the best option based on the limited information that is available. Our focus in this paper is Walter Vincenti's classic case of the Davis wing that was used in the design of the B-24 bomber in World War II (Vincenti 1990). As Vincenti has emphasized, a lack of theoretical knowledge of aerodynamics led to a somewhat open-ended investigation of potential wing designs, focusing on wind-tunnel experiments with small-scale models. These exploratory investigations led the engineers to recommend the use of the Davis wing, despite any clear appreciation of why the wing was working so well. This choice proved quite successful in practical terms as the B-24 bomber became the most widely produced bomber in history (as over 19,000 instances of this type were built). After summarizing this case, we turn to a critical investigation into Vincenti's analysis. Vincenti offers an unsatisfying proposal that the engineers extended their knowledge of airplane wings simply because their choice of the airplane wing proved practically effective. We argue that a more fine-grained analysis of the cognitive capacities and achievements of these engineers is needed to make sense of this case and similar cases of exploratory design. Our positive proposal is developed in terms of three sorts of epistemic achievement. First, the engineers began their investigations with the capacity to plan and manage the design process itself. This allowed them to set up a series of wind-tunnel experiments and to properly interpret the results of these experiments. Second, when the choice of the Davis wing was made, the engineers increased their understanding of how to make a better airplane. This variety of action-oriented or practical understanding (Bengson 2017) was tied to their grasp of what we call a design model for the B-24 bomber. Third, as the B-24 was manufactured and improved, the engineers achieved a greater understanding of airplane wings themselves. This is a kind of enhanced objectual understanding of the phenomenon or subject-matter of their investigation (Dellsén 2017, 2018). The broader significance of this conclusion is that it is fruitful to reconstruct these sorts of exploratory design investigations in terms of planning capacities and understanding. Scientific understanding has received extended consideration by philosophers of science (De Regt 2017; Elgin 2017; Malfatti 2023). However, we maintain that the kind of theoretical understanding that is most emphasized by philosophers of science should be supplemented by an account of action-oriented or practical understanding. By considering these different kinds of understanding and their relations to one another, we argue that the cognitive achievements of both engineers and scientists can be more

*Speaker

clearly appreciated.

Keywords: understanding, artifact design, exploratory models, engineering

A Philosophical Investigation of Emotion Recognition Technology

Alexandra Prégent * [1]

[1] Universiteit Leiden = Leiden University – Netherlands

The rise of sociotechnical systems and practices has impacted our lives by enmeshing people and technology in webs of information, reshaping social dynamics and expectations. The field of neurotechnology, in particular, has raised concerns about the protection of mental privacy, prompting ethical, legal and political debates. "Neurorights" have emerged as a first attempt to protect against new invasive neurotechnologies. This technology aims to gain access to individuals' thoughts, emotions and other mental states by "monitoring brain activity" or "intervening in brain activity" (Ligthart et al. 2023). A surge in emotion recognition technology (ERT) shows similar access to the 'inner self', promising to reveal people's inner emotions, mostly through the use of either electroencephalography (EEG) or sensory data.
While ERTs based on EEG signals appear to be protected by ongoing work on 'neurorights', ERTs based on sensory data do not appear to be covered by these types of rights, leaving their development at greater risk of misuse and associated harm. As emotions play an important role in everyday life - from human interaction to decision making - the normative issues surrounding the use of neurotechnology may now also apply to ERTs based on sensory data. However, while neurorights focus on potential privacy rights over our 'brain activity', sensory data-based ERT raise questions about potential privacy rights over our 'emotional expressions'. Thus, at the intersection of both neurorights and public privacy conundrums, emotional expressions are a particularly challenging and interesting case.

This paper proposes to explore new avenues for potential privacy rights over our emotional expressions by addressing some of the challenges posed by ERTs. It will do so by following an empirical philosophical methodology, which means that the proposed philosophical privacy approach builds on the scientific output provided by affective computer science and social psychology research. Moreover, the privacy approach is rooted in two different social privacy perspectives, 1) Altman's (1975) account and 2) Nissenbaum's (2010) account, revisiting both accounts and proposing a new one that take inspiration on them.
The prospective outcome of this new approach to the problem aims to bridge the gap between natural science and humanities research on ERTs, which it does by integrating and bringing together recent technological developments and revisiting philosophical perspectives on privacy in order to provide pragmatic insights into the use and regulation of this emerging technology.

Keywords: Privacy, Emotion Recognition Technology, Emotional Expressions, Sensory Data, Empirical Philosophy

*Speaker

The ethics and moral psychology of vaccination

Alejandra Petino Zappala [1], Andrea Quint * [1,2], Phuc Nguyen * [1,3], Nora Heinzelmann * † [3,4]

[1] German Cancer Research Center – Germany
[2] Department of Psychology, University of Heidelberg – Germany
[3] Institute of Philosophy, University of Erlangen-Nuremberg – Germany
[4] Department of Philosophy New York University – United States

Even before the COVID-19 pandemic, the World Health Organization named vaccine hesitancy as one of the top 10 threats to global health. We argue and show that moral philosophy and philosophical moral psychology can contribute to responses to this threat. Methodologically, philosophy may thus not only inform the medical and health sciences but also yield real-world benefits.
Societal debates about vaccines have become increasingly moralized. To at least some extent, moral concerns voiced in those debates map onto genuine ethical issues that philosophical theory discusses, such as the under what conditions vaccine mandates are justified. Identifying and addressing genuine moral concerns and clarifying misunderstandings about ethical issues may thus advance not only lay understanding of vaccine ethics but also contribute to the fight against vaccine hesitancy. From the literature in applied ethics, we identified relevant ethical issues surrounding vaccines. Through re-iterated pre-testing in pilot studies, we then developed a questionnaire that translates ethical issues into items accessible to laypersons. We then employed this questionnaire in a pre-registered, cross-cultural survey to identify ethical issues that may affect vaccination intention. We focused on vaccinations against cancer and other diseases that are caused by human papillomaviruses (HPV). To assess how ethical concerns compare and relate to other factors known to affect vaccination intention, we also collected data on these, such as knowledge and trust.

We found that ethical concerns in laypersons broadly align with ethical theory and strongly correlate with vaccination intention. We also found that vaccination intention is stronger the smaller risks of vaccine side-effects and the greater risks of disease are perceived, in line with a utilitarian harm-benefit approach. Trust in doctors, science, and healthcare institutions mediated the negative relation between perceived risks and vaccination intention. Interestingly, effects of knowledge were much smaller than those we found for ethical judgments. This may indicate that campaigns against vaccine hesitancy may be more effective if they not only target lack of knowledge but also highlight ethical aspects of vaccination, such as individuals' concern for the greater good.

*Speaker
†Corresponding author: nora.heinzelmann@fau.de

We are currently conducting a second study measures how possible interventions may affect ethical judgements and vaccination intention. For example, we present two groups of participants with basic information about vaccination and disease; one of them is also shown a peer message appealing to individuals' concern for the greater good. We then compare ethical judgements and vaccination intention of both groups. Initial results from a pilot performed in the US and the UK indicate that UK participants viewing the peer message are significantly more likely to rate HPV vaccination as ethically better and medical institutions as more trustworthy than participants from the other groups.
In sum, this work seeks to demonstrate that philosophical ethics may not only inform the medical and public health sciences but also contribute to real-world change.

Keywords: Philosophy of medical and health sciences, philosophy of medicine, medical ethics, moral psychology, vaccination

Explications in Mathematics

Jonas Raab *† [1], Deniz Sarikaya * ‡ [2]

[1] Ludwig Maximilian University [Munich] = Ludwig Maximilians Universität München – Germany
[2] Vrije Universiteit Brussel [Bruxelles] – Belgium

Carnap introduced his notion of explication to arrive with concepts that are precise enough to achieve scientific progress. To make concepts more precise, it is necessary to start from less precise ones. However, this understanding of explication implies that is not applicable to an important domain: mathematics. We show that within mathematics explications are both possible and more widespread than it might first appear. In particular, we argue that (1) actual mathematical practice explicates; we discuss, examples like topological spaces (which might have been introduced to capture what we now call Hausdorff Spaces), prime numbers (some of whose definitions imply that the number 1 is a prime number and so complicate the prime-factor theorem), and the famous case study of the polyhedron by Imre Lakatos, whose definitions evolve simply to rule out some unintended counter examples. We also argue that (2) such explications are needed for the more modern more precise mathematical endeavour which takes foundational work more serious, namely in automated theorem proving for which it is necessary to expand functions and introduce new explicit conventions which are usually kept rather ambiguous in current practice. For example, there are reasons to make the division total by including division by zero. There also are foundational aspects, that might make some type system desirable, which then enforce the distinction between addition on the reals and on the complex numbers. These cases might appear very basic, but they have ripple effects that either render advanced topics false or enforce being explicit about several of their implicit assumptions.

While this appears to be mostly about uninteresting fringe cases, we argue that this is an interesting development that correlates with arguments of *productive ambiguity* within mathematics. Carnap sees potential explications as *proposals* which are not themselves truth-apt. Rather, we have to find different ways of evaluating the success of explications for which Carnap introduces *adequacy criteria*. It seems that, since there is no 'clear-cut answer', several proposals are possible, and, therefore, several proposals can be *adequate*. Indeed, if only one proposal were possible, then the problem would have already been formulated in exact terms and, therefore, explication would not be possible.

We finally argue that these mathematical case studies also motivate more work on the notion of explication, and have the potential to further close the methodological gap between philosophy and mathematics.

*Speaker
†Corresponding author: JoRaab@gmail.com
‡Corresponding author: deniz.sarikaya@vub.be

Keywords: Explication, Mathematics, Mathematical Practice, Conceptual Engineering

Science Journalism as an Institution of Epistemic Vigilance: The Potential for New Intermediaries

Tanja Rechnitzer * [1]

[1] Leibniz Universität Hannover=Leibniz University Hannover – Germany

Today's communication environment is characterized by information overload and epistemic pollution, making it challenging for laypeople to identify trustworthy experts and scientific information that is both reliable and relevant for their epistemic and practical needs. Science journalism has the potential to help alleviate this problem by serving as an institution of epistemic vigilance and acting as a gatekeeper towards the public for identifying trustworthy science. However, delegating the tasks of identifying reliable and relevant scientific information to science journalism requires that it itself be trustworthy.
This talk explores two interrelated questions: First, what are epistemic virtues and characteristics that science journalism needs to have to be able to fulfill the role of a trustworthy institution of epistemic vigilance? Second, can science journalism adapt to challenges arising from the disintegration of the traditional media-landscape, while maintaining or even strengthening its role as institution of epistemic vigilance?

Addressing the first question, I argue that science journalism needs to exhibit a critical stance towards science, a particular set of skills and competences, as well as a systematic cooperation and structured division of cognitive labor both within science journalism as well as between science and journalism. By scrutinizing and contextualizing scientific studies and scientific expert statements, science journalism can crucially contribute to the public trustworthiness of science.

However, this puts pressure on the collaboration and division of cognitive labor: Scientists need to trust that journalists will only raise legitimate criticism, and that they will be able and willing to correctly represent expert statements. Audiences need to trust that journalists are not only competent in identifying trustworthy experts, but also able and willing to present their findings in a way that is both objective and sensitive to the needs of the audiences. Consequently, science journalism as an institution needs an ambitious set of skills and competences, exceeding those of a single journalist. Thus, the work of epistemic vigilance needs a systematic organization, e.g., through specialization on different areas of competence

To address the second question, I use the example of the Science Media Center Germany (SMC) to discuss how new intermediary institutions could help to update the practices of epistemic vigilance. Science journalism is currently under pressure due to several factors. On the one hand, these concern the changing media landscape, like the 24/7 news-cycle, financial cuts, and the loss of traditional gatekeeper functions. On the other hand, there is an increased epistemic

*Speaker

vulnerability to epistemic failures in science, due to science-internal quality-control no longer working properly, as, e.g., the replication crisis shows.
I show how the SMC helps to alleviate some of these problems by functioning as an organization that identifies experts, contextualizes study results, and provides meta-expertise for journalists without itself producing narrative journalistic articles. It thereby can serve as a broker of trust between scientists and lay audiences. To conclude, I argue that new, independently funded, intermediary organizations are one viable option for updating processes of epistemic vigilance.

Keywords: science journalism, epistemic vigilance, trust, science communication, Science Media Center, intermediary organization

The Rationality of the Sense of Touch: Dialectical Reflections on the Molyneux-Question

Aimen Remida * [1]

[1] Philipps Universität Marburg – Germany

The Molyneux-Question refers to the situation of a congenitally blind person, who used to recognize objects like cubes and spheres by touching them, and to whom the sight is now restored: will he be able to distinguish between a cube and a sphere put on a table in front of him, by merely using his sight and before he touches them? This question was intensively discussed in the eighteenth century and the different answers given by the philosophers can be divided into two general groups: while the empiricists give a negative answer, the rationalists tend to formulate an affirmative one. The Molyneux-question still represents a challenge for today's philosophers, even though the last three centuries have witnessed the elaboration of various approaches and attempts to cope with the difficulties of the problem. In this paper I claim that an appropriate solution of the Molyneux-problem should start with investigating the specificities of the (hitherto neglected) sense of touch, both in its general status within philosophy of perception and in the particular case of the cognition process among blind people. In fact, an analysis of the consequences of relying primarily on the sense of touch in acquiring knowledge can be quite informative regarding the much-discussed relations between sensory experience and concept formation. After a first introductory section, in which the differences between the empiricist and the rationalist accounts are highlighted, I present the outline of a dialectical suggestion that goes beyond the yes-no-dichotomy of the classical answers to the Molyneux-question. This approach argues that perceptual experience has a conceptual content that the subject memorizes in association with further elements such as testimonial beliefs. At the core of the argument lies a suggested connection between empirical evidence and rational inference. This connection rests upon a determined account of the role of memory, which I develop on the basis of selected historical references.

Keywords: Molyneux, question, touch, empiricism, rationalism, dialectic

*Speaker

Hegselmann–Krause Model with Truth Parameters: Network Effects, Opinion Clusters and Truth Approximation

Filippo Riscica * [1]

[1] Alma Mater Studiorum Università di Bologna = University of Bologna – Italy

The Hegselmann–Krause (HK) model with truth parameters (Hegselmann and Krause, 2006) is one of the most influential non-Bayesian agent-based models in formal social epistemology. In an HK model with truth parameters, agents' opinions are influenced by opinions that are sufficiently similar to their own. That is, those opinions that fall within a predetermined distance threshold (called confidence bound). Moreover, some agents, called truth-seekers, consider an objective source of evidence (e.g., an experimental reading) that gives reliable information about the true state of the world. The HK model provides a flexible framework for studying communities of agents adhering to different epistemic norms.
Notwithstanding the influence of the HK model in the literature, there is an important gap with respect to the network effects. Douven and Hegselmann (2022) is the only work addressing network effects, but it suffers from several methodological shortcomings, such as inadequate consideration of randomly generated network structures. Furthermore, Douven and Hegselmann focus on applying the HK model to the bandit problem without first addressing network effects on HK, making it difficult to evaluate whether their findings are attributable to the model's features or the specific bandit problem context.

In this talk, I present the results of the first systematic study of the network effects on the HK model with truth parameters and assess their implications for social epistemology. My analysis is based on Monte Carlo simulations on different network structures (both deterministic and randomly generated) and for different population sizes, providing robust insights into the behaviour of the model. The network structures considered are complete, cycle, cycle with random edges, wheel, and Erdös–Rényi.

I examine how different network structures influence the group's tendency to reach consensus, polarization, or fragmentation (measured by the number of opinion clusters), and how closely the group's opinions approximate the truth (percentage of agents adopting the true opinion). Moreover, I introduce in the study of HK model with truth parameters Shannon entropy as a way to provide a finer-grained analysis of agents distribution within opinion clusters.
My work shows the following epistemically interesting results. First, randomly generated networks are the best performing overall in terms of promoting truth approximation. Within deterministic networks, complete networks outperform cycles and wheels in terms of promoting truth approximation. Thus, my results suggest that the Zollman effect-a phenomenon where less connected networks can sometimes lead to better epistemic outcomes-is not observed in the HK

*Speaker

model. Second, when truth-seekers rapidly update their opinions upon learning new objective information, the group is more likely to end in a state of polarization or fragmentation. Therefore, my results imply that in complex epistemic environments where diverse epistemic norms are present and achieving broad societal consensus is important, groups attain better epistemic states when truth-seekers update their opinions more gradually.

Keywords: Agent Based Models, Epistemic Communities, Network Epistemology, Monte Carlo Simulations, Shannon Entropy

What is effective metaphysics?

Sébastien Rivat * [1]

[1] Munich Center for Mathematical Philosophy (MCMP), LMU Munich – Germany

Metaphysicians of science have been fairly virulent toward contemporary analytic metaphysics during the last decades (e.g., Ladyman et al., 2007). In particular, they have highlighted its tendency to make claims that are in plain contradiction with contemporary physics. For metaphysics to become once again a respectable form of inquiry, they have thus not failed to insist on the necessity of "naturalizing" it, in the broad sense that it must be informed and guided-if not constrained-by science.
In practice, naturalized metaphysics seems to have quickly fallen into similar anti-naturalistic traps, at least in the context of physics. Much of current metaphysical work indeed interprets existing physical theories, such as non-relativistic quantum mechanics and classical general relativity, under the fiction that they are complete and fundamental. But this interpretative practice is blatantly anti-naturalistic: most physicists indeed consider that our best current theories are best understood and formulated as "effective theories", that is, as theories that are, by virtue of their own mathematical structure, not even able to be complete and fundamental. And this interpretive practice also tends to sever the link between metaphysics and epistemology by generating far-fetched metaphysical images that are largely incompatible with one another (e.g., the debate over wave function realism and the primitive ontology approach).

Naturalized metaphysicians thus seem to have every reason to move toward a kind of "effective metaphysics" by setting themselves to the more modest task of uncovering the non-fundamental structure of certain parts or aspects of the actual world, by analogy with "effective physics". But this raises fundamental challenges: (i) How, exactly, should we understand the project of effective metaphysics? Is effective metaphysics even possible (in the sense of delivering metaphysical knowledge)?

My goal in this talk is to respond to Kerry McKenzie's (forthcoming) recent argument against the possibility of effective metaphysics and outline a refined account of what effective metaphysics consists in. I will first briefly present her account. In particular, she argues that the metaphysical content of non-fundamental theories does not approximate in any sense the metaphysical content of a fundamental theory, and I will explain why she takes this to undermine the possibility of effective metaphysics. Then, I will present what I take to be the most decisive worry against her account: namely, that there are decisive examples in physics where the metaphysical content of a non-fundamental theory is largely underdetermined by that of a more fundamental theory (we may speak of "metaphysical multi-realizability" in such cases). Finally, I will suggest that effective metaphysics is best understood as the study of non-fundamental and autonomous items, in their own terms and independently of fundamental metaphysics, where, by 'autonomous', I mean that these items depend only little on fundamental items from the perspective of the physical theories that describe them.

*Speaker

Ladyman J. *et al.*, *Every thing must go: Metaphysics naturalized*, Oxford, Oxford University Press, 2007.
McKenzie K., "On the prospects of an effective metaphysics", *in* F. A. Müller (ed.), *The structure of reality and the reality of structure*, Dordrecht, Springer, forthcoming.

Keywords: Effective Metaphysics, Effective Theories, Naturalized metaphysics, Metaphysics of Science

Why computational complexity may set impenetrable barriers for epistemic reduction

Christian Sachse * [1]

[1] University of Lausanne – Switzerland

The reductionist program in the philosophy of science frequently assumes that the laws and regularities of special sciences, such as biology and psychology, can ultimately be reduced to more fundamental physical laws. This assumption is heavily challenged by the multiple realization argument, which holds that higher-level properties can be realized by various distinct lower-level physical states. In this paper, we introduce a novel argument against epistemic reductionism rooted in computational complexity. We contend that under certain conditions, "complexity barriers" arise, rendering the reduction of higher-level explanations to lower-level physical descriptions impossible, even in principle, assuming physicalism is true.
Drawing on examples from cryptography and complexity theory, we demonstrate how these barriers prevent the discovery of lawful relationships between different levels of scientific explanation. Such barriers impose impenetrable limits on the reductionist approach, suggesting that some scientific phenomena may be inherently irreducible due to their complex nature. This has profound implications for our understanding of the special sciences and their relationship to fundamental physics.

Central to our argument is the idea that certain mappings between levels of explanation, such as from neural states to clinical outcomes, may involve functions that are inherently unlearnable, even with extensive data. These "complexity barriers" function similarly to cryptographic systems, where information, though fully observable, remains undecodable without a specific key. For instance, the inability to predict the severity of Huntington's disease solely from neurobiological data, despite clear genetic correlations, illustrates that complete knowledge of lower-level processes may not suffice to reduce higher-level phenomena. This challenges the foundation of reductionist approaches in the special sciences, particularly when mechanistic explanations lose their explanatory power due to the inability to identify causally specific parts within complex systems.
Finally, we explore the interplay between the multiple realization argument and the complexity argument, highlighting them as distinct yet complementary obstacles to epistemic reduction. While multiple realization suggests that higher-level properties can be instantiated by a diverse array of lower-level states, complicating efforts to reduce these properties to a single physical description, the complexity argument advances this by proposing that even a seemingly possible reduction may be obstructed by insurmountable computational barriers, as no realizer types may be identified. Together, these arguments present a nuanced critique of reductionism, suggesting that the challenges to reducing special science theories to physical theories are both profound

*Speaker

and varied. This discussion not only deepens our understanding of the limitations of reductionist approaches but also opens new avenues for philosophical exploration and scientific methodology, particularly in investigating how complexity barriers influence scientific progress.

Keywords: reduction, complexity, multiple realization, reductive explanations

What does constitutive mean?

Fasol Samuele * [1]

[1] Università degli Studi Roma Tre = Roma Tre University – Italy

Over the past two decades, numerous authors have identified specific elements of scientific theories as 'constitutive', without adequately defining this notion. In this talk, I aim to elucidate the various meanings of this term, tracing it back to its Kantian roots, to then compare it with contemporary frameworks concerning specific principles of physics. More specifically, I defend three claims.
First, I claim that Kant used the term "*konstitutiv*" both in a *narrow* and *broad* sense, referring respectively to intuition and experience. This point is deepened and advanced within a framework that does not strongly commit Kant's transcendental philosophy to Newtonian mechanics.

Second, by examining both the Analytic of Principles and the Appendix to the Transcendental Dialectic in the *Critique of Pure Reason*, I claim that constitutive principles *functionally define* what qualifies and what does not qualify as an object of intuition and as an object of experience. Thereby, they endow their respective target to be included in meaningful statements (Kant's judgments), i.e. statements exhibiting a truth value. This activity of constitution transforms the notion of 'object' from an independent and merely present substance to a set of functionally related values. Furthermore, the previous analysis enables comparing constitutive principles with the formal notion of 'presupposition' and sharply differentiating them from other similar notions, like 'premise', 'axiom', 'necessary condition', and 'convention', due to their unique semantic role.
Third, I defend the thesis that constitutive principles are indissolubly intertwined with transcendental arguments. Indeed, transcendental arguments justify the modal status of such principles, setting them apart from empirical, contingent, and non-necessary principles. These criteria may prove useful in evaluating the various post-Kantian reformulations of constitutive principles. Moreover, my analysis distinguishes the constitutive from the regulative principles addressing also the fact that they cooperate and merge in the empirical investigation of nature.

Keywords: Constitutive, physical principles, Kant, Friedman

*Speaker

ML approaches in personalized psychiatry and the idea of hypothesis-free research

Johanna Sarisoy * [1]

[1] University of Edinburgh – United Kingdom

Recent advancements in personalized psychiatry have been significantly bolstered by machine learning (ML) approaches, fueled (amongst other things) by the increasing availability and variety of patient data. Vast datasets are now available with online social media data, mobile sensor data, and extensive collections of brain structure, function, genetic, and behavioral data by different consortia and funding bodies (e.g., ENIGMA, ABIDE, ADNI). Traditionally, ML methods have emphasized predictive capabilities over hypothesis testing. Key objectives in this field include identifying new biomarkers, stratifying patient groups based on symptoms or genetic profiles, and discerning subgroups based on treatment responses.
As personalized psychiatry enters this new era, debates arise regarding the epistemic and scientific advantages of ML-based predictive approaches compared to classical approaches. A prominent discussion revolves around the notion of "hypothesis-free" ML approaches and their perceived benefits over traditional hypothesis-testing methods. Advocates of these methods argue that ML's capacity to operate without predefined hypotheses provides several epistemic advantages. Proponents suggest, for example, that ML approaches have the potential to uncover unexpected patterns, reduce biases associated with human and theoretical assumptions, and allow the data to reveal insights independently of existing theories. Proponents suggest that because predictive ML methods are not constrained by preconceived hypotheses, they enable a more objective exploration of data, offering insights that are less influenced by human bias.

In contrast, I argue against the notion that ML approaches are genuinely hypothesis-free. ML techniques, while not explicitly testing hypotheses, are guided by several implicit assumptions. I offer several arguments to support this thesis: First, ML methodologies operate under the hypothesis that psychiatric phenomena can be understood as complex systems or network dysfunctions. Furthermore, the choice of ML algorithms influences the types of patterns and structures that can be detected, implicitly assuming specific structures within the variable space. Additionally, the selection of training data reflects covert hypotheses. The selection of the data space reflects what type of data is considered theoretically relevant, practically useful, or safe. Moreover, the training data inevitably reflects foundational hypotheses about psychiatric conditions (e.g. as disorders) or concepts, such as 'anxiety'.
The presence of implicit assumptions within ML approaches challenges the claim that such methods are entirely free from theoretical influences or biases. Instead, the structure of the algorithms and the choice of training data both introduce and perpetuate underlying assumptions, which can shape the findings and interpretations of ML analyses. As a result, rather than being a neutral or purely objective tool, ML is inherently influenced by the theoretical and practical frameworks that inform its design and implementation. Recognizing these underlying

*Speaker

assumptions is crucial for a more accurate and critical evaluation of ML's role in advancing personalized psychiatry and understanding its contributions to scientific knowledge.

Keywords: hypothesis, testing, machine learning, psychiatry, value, laden ideal

AI and the discovery of new things

Samuel Schindler * [1]

[1] Aarhus University [Aarhus] – Denmark

Recent years have witnessed the emergence of AI-assisted discoveries, such as the solution of the protein folding problem by Google's AlphaFold. But AI has yet to discover things that are still unknown to us.
AlphaFold is said to have solved the 50-year-old protein folding problem, i.e., it successfully predicts protein structures from protein sequences. The successes of AlphaFold are based on supervised learning, i.e., AlphaFold is a neural network that is trained to predict known protein structures from known protein sequences. Once so trained, AlphaFold can then predict as of yet unknown protein structures from known protein sequences. However, there has never been any doubt that what AlphaFold predicts are indeed protein structures. But can AI make discoveries of new objects whose identity is yet unknown to us?

For example, oxygen was entirely unknown to Priestley when he first discovered it. The theoretical framework he was working within did not allow for such a thing. He therefore had to invent the new category of "dephlogisticated air". And of course, the discovery of oxygen ultimately led to the overthrow of the phlogiston theory in the hands of Lavoisier.

Many other celebrated successes of AI-assisted discovery concern the re-discovery of scientific objects, such as Newton's law of gravitation (Lemos et al. 2023) or the heliocentricity of our planetary system (Iten et al. 2020). However, there have also been efforts to detect entirely new objects: physicists have begun to look for physics beyond the standard model with the help of deep learning methods (Crivellin and Mellado 2024; Farina et al. 2020; Kasieczka et al. 2021; Karagiorgi et al. 2022). These searches use unsupervised learning techniques, i.e., AI is trained without a particular target, allowing for the detection of anomalous events, at least in principle. Boge has recently argued that in these kinds of "exploratory" AI searches, scientists face the challenge of formulating new concepts for understanding the "underlying mechanisms" that is especially difficult given the opacity of AI (Boge 2021). The signals detected by AI-driven searches must be interpreted. And this may be difficult when the things detected by AI are not yet known to us. But this is no different from the need to interpret experimental outcomes. More importantly, the opaqueness of AI may make the task of distinguishing signal from noise not only difficult, but perhaps impossible. Not only in the context of physics, but also elsewhere. This is particularly problematic if, as I'm prepared to argue, truly ground-breaking discoveries require the discovery of yet unknown objects.

Keywords: artificial intelligence, discovery, anomaly searches, artifacts, experiments

*Speaker

On prospective ethics - How to overcome the lack of empirical evidence

Sebastian Schuol * [1]

[1] Friedrich-Alexander Universität Erlangen-Nürnberg = University of Erlangen-Nuremberg – Germany

Historically, a change in perspective can be seen in applied ethics. Previously, solutions were only sought when specific problems arose. However, these solutions often came too late for the case in question. Lessons have been learned from this problem, and now the ethical potential for problems is discussed as soon as a new technology is developed. In this change from a 'retrospective' to a 'prospective' ethics the state of knowledge is epistemically interesting. While in the mode of retrospective ethics one starts from facts, a real problem is to be prevented, in the case of prospective ethics there is (as yet) no such knowledge. Researchers have to rely on speculation and there is no guarantee that the scenario will occur in the future.

For example, in the discourse on the modification of the human germ line one assumes that the unrestricted use of new genetic engineering methods enables a new eugenics. Since this involves selection at the level of genetic makeup, the future of humanity would be affected. In this ethical discourse, the 'uniformity thesis' receives a lot of attention. According to this, decisions concerning reproduction are by no means as diverse as expected in a liberal society, in which the diversity of plays a central role. Rather, it is claimed that, even without state control (classical eugenics), actors are subject to social influences and ultimately act in a similar way. With widespread use of this technology and over time, the genetic diversity of humanity would be reduced, and its gene pool would become uniform. This is be problematic, valuable resources for development would be lost. Should the environmental conditions and thus the challenges for humanity change over time, these genetic deviations would no longer be available for evolutionary development and adaptation would be prevented. However, due to a lack of empirical evidence, the uniformity thesis is just that: a thesis. As speculation about the future, it can neither be confirmed nor refuted.

In my talk, I would like to present a way to overcome this deficiency. In cooperation with a museum an interactive exhibit was developed: On the one hand, the exhibit is intended to inform visitors about the state of knowledge and the ethical challenges of modern genetic engineering (science communication). On the other hand, it is designed to record visitors' preferences regarding possible future applications of genetic engineering so that these can be taken into account in the ethical discourse. Do visitors actually want the same traits (which would confirm the uniformity thesis) or not? And if preferences do emerge, which traits would they choose? Important information for a differentiated and realistic ethical debate about the future of humanity can be gathered. Since empirical data is collected and fed into the primarily theoretical discourse, the project can be methodically located in the narrower context of experimental philosophy. The lecture will present and discuss the method and the first findings from an evaluation of the data.

*Speaker

Keywords: ethics, prospective, exhibit, experimental philosophy

Value Neutrality, Hypothetical Value Statements and Inductive Risk in Science

Gerhard Schurz * [1]

[1] Heinrich Heine Universität Düsseldorf = Heinrich Heine University [Düsseldorf] – Germany

In this talk a new version of the thesis of scientific value neutrality is developed.
In the first part of the talk the value-neutrality thesis will be cleared from misunderstandings and precisely defined as follows: *The justification of scientific knowledge should be independent of fundamental non-epistemic value assumptions.* Although extra-epistemic interests may influence the choice of the research question, it is argued that the fixation of the research question completely screens off extra-epistemic interests from the choice of the epistemically optimal method.

In the second part of the talk it is shown how the value-independence of scientific knowledge can be connected with scientific value engagement: by the method of *hypothetical value statements*. These are conditional value recommendations that are derived by means-end reasoning from hypothetically assumed values, which do not come from scientists but from science users. Two kinds of means-end inferences are distinguished: those based on necessary means and those based on optimal means. Both kinds of means-end inferences presuppose that the benefit of the end that is achieved by implementing the means outweighs the costs of the means. Therefore the premises of means-end inferences are required to be transparent in regard to all side-effects of the means. The idea of value-engagement based on the explained method of value-neutral means-end reasoning is illustrated with experiences from the Covid-19 pandemic.

In the third part of the paper, we attempt to show that the same hypothetical method can also solve the problem of "inductive risk" without having to resort to categorical value judgments. For this purpose we refer to the tragic case of the L'Aquila earthquake. The proponents of the argument from inductive risk (AIR) argue that the qualitative acceptance of a practically relevant but uncertain hypothesis - such as the prediction of an earthquake implying that villages should be evacuated - depends on extra-epistemic cost-benefit considerations. This undoubtedly true. But does this mean that scientists should make practical decisions *on* behalf of the affected people, as proponents of the AIR have proposed? Our answer to this question is *no*. Instead, scientists should explicitly state the involved error risks and make them transparent to non-experts through *hypothetical* cost-benefit consideration. This is called the condition of *uncertainty transparency*.
In the final part of the talk further controversial questions will be resolved through the method of hypothetical value statements, including the "new demarcation problem".

*Speaker

Keywords: values in science, value neutrality, hypothetical value statements, argument from inductive risk, new demarcation problem

From objectual to explanatory understanding with AlphaFold2

Annika Schuster * [1]

[1] Technische Universität Dortmund [Dortmund] – Germany

Deepmind's AlphaFold2 (AF2) deep neural network (DNN) (Jumper et al. 2021) gained a lot of attention when it considerably surpassed other algorithmic devices for protein structure predictions from amino acid sequences at the Critical Assessment of Protein Structure Prediction (CASP). After the first excitement about the results passed, critical voices, however, remarked that the most important questions concerning protein folding are still unanswered (e.g. Outeiral, Nissley, and Deane 2022). After all, science is searching for explanations and aims to enhance *understanding* and *knowledge* of a subject matter, not just highly accurate predictions (e.g. De Regt 2017, Elgin 2017). The black-box nature of DNNs seems to stand in opposition to understanding and knowledge because even understanding the architecture of a successful DNN does not imply an understanding of which features of the input data were responsible for its successful predictions. Indeed, due to the usually vast dimensions of these data and the network architecture, there is no straightforward way to extract this information post hoc.

Philosophers of science have advanced arguments for the possibility of understanding with black-boxes, but also arguments against it. Sullivan (2022) argues that it is not the DNNs' black-box nature that inhibits scientific understanding, but the degree to which they contain link uncertainty, that is, in how far their outputs fail to be connected to the phenomena they are supposed to explain. For example, a DNN that predicts melanoma based on visual data, is so clearly related to the phenomenon and general medical practice that its black-box components are not hindering the possibility to understand how it does what it does, while a DNN that successfully predicts sexual orientation from pictures is not connected to anything known to apply in reality and thus cannot further understanding. Räz and Beisbart (2022) argue contra Sullivan that only "weak" objectual understanding can be gained from DNNs as long as they are not better understood, as deriving actual explanations from DNN predictions would require a much deeper insight into their inner workings.

Building on a case study on the use of AF2 predictions in protein biology, I will show that the influence of objectual understanding provided by DNNs is not to be disregarded as it figures prominently in a new and exciting process that generates scientific explanations and knowledge. AF2 remains a black box in both the sense that its overall functioning is intransparent and that it is unclear how any single prediction comes about (Creel 2020). Yet, its predictions greatly facilitate protein research. As I will argue, an analysis of a great number of protein structures clearly means an increase in objectual understanding in the domain (Bordin et al. 2023). But this is not all – in a second step, AF2's predictions are used to generate explanatory understanding, e.g. when they were used to elucidate structures in highly complex nuclear pore complexes (Mosalaganti et al. 2022).This two-step adaptive process will be presented in my talk.

*Speaker

Keywords: understanding, artificial intelligence

Reliability and Interpretability in Science and Deep Learning

Luigi Scorzato * [1]

[1] Accenture – Switzerland

In recent years, the question of the reliability of Machine Learning (ML) methods has acquired significant importance, and the analysis of the associated uncertainties has motivated a growing amount of research. However, most of these studies have applied standard error analysis to ML models. Is this practice justified? Standard error analysis is routinely applied in science. However, Deep Neural Network (DNN) models represent a rather significant departure from standard scientific modelling. It is therefore necessary to integrate the standard error analysis with a deeper epistemological analysis of the possible differences between DNN models and standard scientific modelling and the possible implications of these differences in the assessment of reliability. This article offers several contributions. First, it emphasises the ubiquitous role of model assumptions (both in ML and traditional science) against the illusion of theory-free science. Secondly, model assumptions are analysed from the point of view of their (epistemic) complexity, which is shown to be language-independent. It is argued that the high epistemic complexity of DNN models hinders the estimate of their reliability and also their prospect of long term progress. Some potential ways forward are suggested. Thirdly, this article identifies the close relation between a model's epistemic complexity and its interpretability, as introduced in the context of responsible AI. This clarifies in which sense-and to what extent-the lack of understanding of a model (black-box problem) impacts its interpretability in a way that is independent of individual skills. It also clarifies how interpretability is a precondition for a plausible assessment of the reliability of any model, which cannot be based on statistical analysis alone.

Keywords: complexity, responsible AI, Epistemic Values, Theory Selection, Scientific Progress.

*Speaker

CANCELLED

Trade-offs and path-dependencies in animal model choice: a case study in xenotransplantation research

Isaïa Seguinot * [1], Simon Lohse [1]

[1] Radboud university [Nijmegen] – Netherlands

In early 2022, the field of organ transplantation achieved a significant milestone with the first successful pig-to-human heart transplant. This breakthrough has been widely publicized, featuring claims that pig-derived organs will solve the organ shortage crisis. While pigs have certainly gained momentum as potential organ donors, they are but one of the many different animal species used in xenotransplantation (XT) research. Such animals serve not only as organ providers, but also as experimental models to investigate and overcome the challenges involved in crossing species boundaries. This raises several important questions: Why have pigs specifically been dominant in the XT discourse? What specific roles do other animal species play, and what does it take for one to be considered promising for translational XT research?
These questions point to a broader discourse on the complexities involved in selecting animal models for biomedical research and the implications of these choices for translating preclinical findings into clinical practice. Research in the history and philosophy of science suggests that the selection of animal models in the life sciences is guided by a multifaceted interplay of factors across different levels. It is not solely a product of scientific considerations, but is also shaped by pragmatic and societal contexts, as well as historical contingencies that create favourable conditions for a particular animal species to become an established model at a given time.
This paper contributes to the discussion on animal model selection by focusing on xenotransplantation, a field often overlooked in the animal modelling literature. This case study sheds light on a field where animal models play a unique role, and enriches the broader discussion on 'good-enough' animal models and the multidimensionality of organismal choice in the life sciences. The paper is structured in two main sections: The first section provides theoretical background on the various kinds of factors influencing animal model choice, emphasizing the trade-offs and synergies involved in such contextualized selection processes. The second section examines the scientific literature on XT, exploring the field's aims, practices, and the roles of different animal models, with a focus on pigs. We discuss how these dynamics create specific path dependencies in XT research and assess whether pigs ultimately qualify as 'good-enough' donors for clinical applications. The conclusion summarizes our findings and reflects on their philosophical and normative implications for the translation of xenografts from 'bench to bedside'.

Keywords: organismal choice, animal experimentation, xenotransplantation, animal modelling

*Speaker

Chemical reactions are productive causal relations

Vanessa Seifert * [1]

[1] National and Kapodistrian University of Athens – Greece

Chemical reactions are processes by which one or more chemical entities transform without changing the atomic number of the relevant chemical elements. Reactions are everywhere and they are a defining feature of chemistry. Yet, in the philosophy of chemistry they have not received much attention as it is implicitly assumed that an understanding of the nature of chemical entities suffices to understand how they transform and interact with each other. However, chemical reactions are often spelled out using terms that are suggestive of causation, thus prompting a discussion of whether there is something more to them that warrants philosophical analysis. To this end, I examine whether chemical reactions could be understood as causal relations and argue that empirical evidence best supports viewing them as causal relations of the productive type.

To support this view, I examine what are the relata of this putative relation. While the reactants of a chemical reaction are uncontroversially part of its cause, it is not clear whether one should also warrant the thermodynamic conditions or catalysis as partly causing it. In chemical practice neither are considered as substantively causing a reaction. Nevertheless, both play an ineliminable role in its realisation. Specifically, a change in thermodynamic conditions can lead to different products, even when the reactants are the same. Secondly, a catalyst substantively takes part in the process of a reaction even though at the end of it, it does not transform into a different entity. This shows that a classic problem of causation arises in the case of chemical reactions: namely how to distinguish background conditions from genuine causes.

To resolve this problem, I employ Ned Hall's distinction between causation as production and causation as dependence. I show that understanding chemical reactions as causal relations of the productive type helps us distinguish genuine causes from background conditions in a manner that is consonant with how thermodynamic conditions and catalysis are standardly understood in chemical practice. To further support the productive view of causation for the case of chemical reactions, I discuss two other features of chemical reactions, namely reaction mechanisms and affinity tables. Based on their analysis, I argue that productive theories of causation capture more faithfully the nature of chemical reactions.

All in all, a productive view of causation is better supported by empirical evidence for the case of chemical reactions. This is a surprising result if we take into account that Humean accounts of causation (under which causation as dependence falls under) are standardly considered as being better supported by what we observe and know through science.

*Speaker

Keywords: Causation, Chemical reactions, Production, Dependence, Philosophy of Chemistry

Creative abduction in the lab: Lessons from early electrophysiology

Maria Serban * [1]

[1] University of East Anglia [Norwich] – United Kingdom

Luigi Galvani's frog leg twitching in response to an electrical spark set in motion a series of experiments that would lay the foundation for modern electrophysiology. Tracing the links between Galvani's frog leg experiments, the development of Alessandro Volta's pile, Emil du Bois-Reymond's precise measurements of bioelectric currents, and Julius Bernstein's membrane theory, this paper questions what drives the creative process in experimental science. By unpacking the problem-solving nature of experimentation, I articulate an iterative model of the inferential strategies underpinning scientific creativity.
The starting hypothesis is that in these early electrophysiological studies, experimentalists were not merely testing pre-formulated hypotheses, but actively exploring a complex problem space, using various inferential strategies to navigate unfamiliar terrain. Key to this exploration seems to be a form of reasoning that has traditionally been labelled "abduction." However, the abductive inferences we observe in these historical cases challenge standard philosophical explanation-based accounts of abduction. In response, I propose that a conditional approach better captures the ways these scientists generated and evaluated hypotheses in their experimental contexts. For instance, in Volta's response to Galvani's "animal electricity" one can notice that rather than simply accepting or rejecting Galvani's explanation, Volta engaged in a creative abductive process, which can be linked to the invention of the Voltaic pile, an experimental design that not only challenged existing theories but opened up new avenues for experimental investigation. A conditional-based account of abduction also allows for a more contextualised representation of the inferential processes involved in experimental problem-solving. I illustrate this claim by showing how scientists like du Bois-Reymond and Bernstein built upon their predecessors' work, each creative abductive inference step reframing the problem space and suggesting new experimental possibilities. Throughout this analysis, I understand creativity as a certain type of successful exploration of the problem space in which the experimentalist operates.

This in turn raises the question of whether other inferential steps (deductive and inductive) also exhibit creative features. Focusing on Bernstein's application of Nernst's equation to cellular membranes, I challenge the standard view that there is no creativity involved in the deductive and inductive stages of experimentation. The investigation of experimental creativity then challenges us to reevaluate the relationship between exploratory and confirmatory experiments, suggesting a more fluid, iterative view of the different stages of the experimental process. In the early electrophysiology cases, I outline the continuum of experimental practices, emphasising where exploratory abductions led to confirmatory designs, and supposedly confirmatory experiments opened up new exploratory paths. For instance, Du Bois-Reymond's precise measurements, intended to confirm existing theories, led to unexpected observations that sparked

*Speaker

new rounds of creative abduction.

I claim that by tracing the early history of electrophysiology through this lens, it is possible to obtain more nuanced philosophical concepts of abductive inference and experimental creativity. The case studies used in this paper actively contribute to reframing experimentation as a creative, socially embedded problem-solving efforts driven by conditional abductive reasoning.

Keywords: experimentation, electrophysiology, creativity, abduction

The Criminalist's Paradox as a Counterexample to the Principle of Total Evidence

Michał Sikorski * , Alexander Gebharter [1]

[1] Marche Polytechnic University – Italy

The Principle of Total Evidence (PTE) posits that all relevant information should be considered when assessing the truth of a hypothesis. In my presentation, I argue that PTE clashes with the Relevance Principle (RP), a methodological principle widely accepted in forensic science, which claims that only task-relevant factors should be used during certain epistemic tasks (see NCFS 2016). Task-relevant factor (e.g., information) changes the probability of collected evidence (e.g., the fingerprint collected at the crime scene) given the hypothesis (e.g., the person of interest is the source of the fingerprint collected at the crime scene) is true. I will present the methodological discussion of bias in forensic science, which provides the context and justification for the RP.
In response to concerns raised in an important report presented by the National Research Council (NRC (2009)), a number of studies were conducted to explore the extent to which task-irrelevant factors (e.g., information concerning the subject's admission of guilt or results of other forensic procedures) can bias forensic analysis and their results. The results of demonstrate that such factors have the potential to change the results of forensic analyses and divert them from truth (see Cooper and Meterko 2019). Consequently, RP was proposed to distinguish biasing and therefore unacceptable influences from task-relevant and acceptable ones.
Then I will present a Criminalist Paradox that constitutes a counterexample against the PTE (see e.g., Thompson, 2010). The paradox arises from the fact that reliance on certain evidence can make the results of forensic procedures more reliable but at the same time dependent on the accommodated evidence, and therefore less useful in a court case. For example, a fingerprint expert can be more reliable if she is informed by the results of DNA analysis conducted on the same trace (DNA analysis is typically much more reliable than fingerprint analysis). However, this influences the results of the fingerprint evidence depending on the results of DNA evidence and therefore makes it less useful as forensic evidence. Evidence presented during court cases has to be independent from other evidence. Consequently, such influences are excluded by RP and avoided in forensic science. At the same time, PTE recommends using some of the biasing information, creating tension between the two principles. Finally, I will argue that the motivation for adopting RP generalizes beyond forensic science, discuss some possible objections to the counterexample, and propose a new version of PTE that is not susceptible to it.

Keywords: Principle of Total Evidence, Bias, Criminalist's Paradox

*Speaker

Developing A Physics Specific Theory of Analogy: A Case Study of Black Holes and Sonic Black Holes

Veljko Simovic *† [1]

[1] University of Western Ontario – Canada

Analogical arguments have been indispensable in the progress of science. This raises a question: What key features make an analogical argument fruitful i.e., make it lead to new experiments, insights, or explanations? Hesse (1966) suggests it is (a) the **observability** of properties of the alleged analogs i.e., properties of the domains under comparison, (b) the **causal connectedness** of the properties within the domains, and (c) the **absence of essential difference** between the domains. As Hesse aims to capture the key features of all (potentially) fruitful analogical arguments in science, she keeps the characterization of these features general. However, as Norton (2018) argues, Hesse's already general characterization of the key features is bound to become increasingly general. This is because the only way for Hesse's characterization to avoid being too restrictive, is for it to go through infinitely many rounds of embellishments. To keep Hesse's strive and address Norton's worry, I propose a physics-specific theory of analogy.
The focus on physics enables me to offer a more specific characterization of Hesse's key features and thus dull the edge of Norton's worry. Inspired by physical practice, I specify Hesse's observability by adopting Chang's quality-based approach (2005); I specify her causal connectedness by adopting Reichenbach's probabilistic causality (1956) and I specify her 'absence of essential difference' by linking it to observability. Furthermore, I suggest an additional feature – robustness (Schupbach 2018) – which I take to be a strong indicator of the persistence of a property through scientific change. I contend that if an analogical argument in physics possesses all these features, one should be supremely confident in its fruitfulness. This is unlike Hesse, who argued that an argument possessing the key features as she defined them, must be fruitful. This weakening of Hesse's aspiration, I argue, can serve to further ameliorate Norton's worry.
To demonstrate the usefulness of my theory, I apply it to the analogy between black holes and sonic black holes. (Visser 1999, 2011) The assessment shows that although the analogy is overall promising, it has two weak points: 1) the Hawking radiation in the astrophysical scenario is arguably unobservable and thus cannot be judged for robustness; 2) Kerr and Schwarzschild metrics used to describe black holes, and the acoustic metric used to describe sonic black holes, despite their mathematical similarity, seem to fundamentally differ in their physical interpretations due to different variables that figure in them. In answer to the first problem, one might argue that there is nothing in principle preventing the observability of Hawking radiation. Also, the trans-Planckian problem related to the immeasurability of Hawking radiation has its parallel in fluid mechanics where a similar breakdown occurs due to the neglected atomic nature of the fluid. (Dardashti, Thebault and Winsberg 2014) As for the second problem, the introduction of gravity to the acoustic metric suggested by Alibabaei (forthcoming), might help bridge the

*Speaker

†Corresponding author: vsimovi@uwo.ca

gap between it and its astrophysical counterparts. If the above weaknesses are addressed, one should have supreme confidence in the fruitfulness of the analogy.

Keywords: analogy, black holes, sonic black holes, Hesse, Norton

CANCELLED

Substantiating Downward Causation with Production

Kaamesh Singam * [1]

[1] Indian Institute of Technology [Kanpur] – India

James Woodward in his articles (2020), (2021) argues that downward causation is present between the membrane potential (represented by V) and ionic conductances (represented by gK and gNa) in the Hodgkin-Huxley model of the action potential. This presentation aims to challenge the adequacy of the notion of downward causation presented by Woodward and in turn provide additional conditions for the same. First, a notion of levels, following (Povich & Craver, 2018) and (Gillett, 2021), is developed to categorise the different variables (or whatever in the world that the variables represent) into various levels within the mechanism of the action potential. Then, it is argued that the membrane potential V and the conductances gK and gNa are not really at different levels as Woodward supposes them to be. While thin causation is present between V and gK and gNa, in the sense that the relationship counts as causal within Woodward's interventionist framework, there is no evidence to suggest that the relationship is downward causal. If anything, Woodward's notion of conditional independence shows that the above variables are all at the same higher level. Thus, it is concluded that there is no downward causation between the membrane potential and the ionic conductances. From here, new conditions are proposed to strengthen the existing account of downward causation. Making the distinction between thin and thick causation, following (Hall, 2004), it is argued that variables (or whatever in the world that they represent) must have thick causal (productive) relations among them if they are to be considered to belong at the same level. In the case of the HH model of the action potential, even though the membrane potential and the ionic conductances are variables that exhibit interventionist dependence relations, they do not exhibit production between them. The above claim is shown to be true through a simplified account of production. While productive relations exist between the ions (and their electric fields) and the individual ion channels (and their structural configurations), they don't seem to exist between V and gK and gNa, which are the variables generated through coarse-graining from the electric fields and the conductances of the individual potassium and sodium ion channels. This is corroborated by the fact that even the manipulations introduced to the higher-level variables are realised only through the changes in the values of lower-level variables. Following Gillett (2016) and (2017), it is argued that downward causation cannot exist between a whole and its own parts; it can only be diagonal and hence can only exist between a whole at a higher level and the parts of another whole at a lower level. However, Gillett's requirement of a downward (non-causal) determination from a whole to its own parts is relaxed and, instead, the requirement of production to exist between the relevant wholes at the higher level is enforced. Thus, it is argued that downward causation can only exist between levels where production exists. Following this account, some other potential examples of downward causation are analysed.

*Speaker

Keywords: downward causation, levels, production, mechanism, HH model, action potential, conditional independence

Inventing Risk: Ramsey and de Finetti on risk/uncertainty distinction

Caterina Sisti *† [1], Luca Zanetti * ‡ [2]

[1] Scuola Normale Superiore di Pisa – Italy
[2] School of Advanced Studies (Scuola Universitaria Superiore) (IUSS) – Italy

Economists have distinguished between risk and uncertainty since the early 20th century: 'risk' refers to situations in which the probabilities of the outcomes of one's actions are known, whereas 'uncertainty' refers to cases where the outcomes or their probabilities are unknown. A distinction between numerical and non-numerical probabilities was introduced by Keynes in A Treatise on Probability (1921), which can be interpreted as corresponding to risk and uncertainty, respectively. In Risk, Uncertainty, and Profit (1921), Knight argued that 'uncertainty must be taken in a sense radically distinct from the familiar notion of Risk ... a measurable uncertainty, or "risk" proper, ... is not in effect an uncertainty at all'. Keynes later clarified that 'the sense in which I am using the term ("uncertainty") is that in which ... there is no scientific basis on which to form any calculable probability whatever. We simply do not know' ('The General Theory of Employment', 1937).

The distinction between numerical and non-numerical probabilities, or risk and uncertainty, as articulated by Keynes and Knight respectively, is often regarded as irrelevant from the perspective of the subjectivist interpretation of probability. According to the subjectivist interpretation, probabilities are degrees of belief. As such, subjective probabilities can always be measured. This point was independently made by Ramsey in 'Truth and Probability' (1926) and by de Finetti in 'Probabilism' (1931). In Economia delle Assicurazioni ((Economics of Insurance), 1967, with Emanuelli), however, de Finetti offered a more positive assessment of Knight's distinction. According to de Finetti, the distinction between risk and uncertainty is not sharp. On the contrary, 'risks are those that involve the least discrepancy between individual probabilities' whereas uncertainty is characterized by disagreement between individuals in their subjective estimates. Apparently, no similar discussion can be found in Ramsey's work.

The goal of this paper is to offer a partial unification of Ramsey's and de Finetti's views. Specifically, we propose an interpretation of Ramsey's statistical generalisations that aligns his ideas with de Finetti's intersubjective perspective. To achieve this, we draw on Ramsey's account of open generalisations, which he calls 'variable hypotheticals'. In 'General Propositions and Causality' (1929), Ramsey defines variable hypotheticals as generalisations that range over an infinite domain. They are 'rules for judging' that guide agents in forming new beliefs. In 'Chance' (1928), Ramsey claims that such generalisations are a kind of objective degree of belief from which an agent's actual degrees of belief are derived. However, variable hypotheticals depend on an agent's personal experience and, as such, they may differ among agents. We argue that this subjective element in the characterisation of generalisations can be used to recover a distinction

*Speaker
†Corresponding author: caterina.sisti@sns.it
‡Corresponding author: luca.zanetti@iusspavia.it

between risk and uncertainty similar to the one proposed by de Finetti. In turn, this interpretation can shed light on Ramsey's subjectivist position. We conclude that Ramsey's view can supply de Finetti's perspective with a stronger philosophical background.

Keywords: Uncertainty, Bruno de Finetti, Frank Ramsey, Subjective probability, Variable hypotheticals

Distributive Epistemic Injustice in Climate Modeling

Maria Sojka * [1]

[1] Heinrich-Heine-Universität Düsseldorf – Germany

The role of values in science has traditionally often been discussed as a primarily epistemic problem, with the potential for some entailing ethically concerning consequences. A central question of philosophical debates about the adequacy of non-epistemic values in science has been the extent to which these values lead scientists to draw epistemically problematic conclusions. This has also been a particular focus of discussions about the significance of non-epistemic values in climate modeling. However, as research on the role of non-epistemic values in climate modeling further has shown, there are instances where non-epistemic value judgments in intra-scientific processes can be, at the same time, unavoidable, epistemically harmless and ethically problematic. That is, they do not lead to the endorsement of 'false facts' but have nevertheless undesirable consequences. This is due to the complexity of the climate system, which requires scientists to make decisions regarding which parts of the climate system are modeled with more accuracy than others. Thus, value-laden judgments about the extent to which particular processes and variables are included in the models have to be made. Climate modeling is a protracted and expensive undertaking, and modeling centers are predominantly located in the Northern Hemisphere, particularly in Europe and North America. Therefore, there is a risk that models disproportionately reflect the epistemic interests of those who created and paid for the models. More generally speaking, there is a risk that the epistemic needs of particular marginalized communities for specific kinds of information are overlooked, impacting their ability to make policy decisions. I argue that when these communities are not in a position to generate this knowledge on their own, it can lead to distributive epistemic injustice. Models as complex as climate models are particularly prone to distributive epistemic injustice, as they are so expensive that certain groups must rely on others to consider their predictive needs. Further, I argue that the unique epistemic challenges of complex computer simulations, particularly their epistemic opacity, can make it a difficult if not unattainable task to determine whether distributive epistemic injustice is actually an unintended feature of these models. While it is often argued that scientists should make their reasoning process public to counteract the unavoidability of value-laden decision making in science, this is not sufficient in the context of climate modeling. Due to epistemic opacity of climate models and the vast array of epistemically not fully constrained decisions done over the many years of their development, it is often practically impossible to determine the effects specific choices have on the overall model. I conclude that this also affects our ability to assess the consequences of particular modeling decisions for the epistemic needs of specific communities. It further impacts the extent to which we can rely on other possible countermeasures, such as increasing the social diversity among researchers or taking into account the standpoints of marginalized groups, to prevent epistemic inequalities.

*Speaker

Therefore, I argue the complexity of climate modeling can give rise to a particularly intricate form of distributive epistemic injustice.

Keywords: Climate Modeling, Epistemic Injustice, Values in Science

Communicating Uncertainty: A Metacognitive Approach

Sergiu Spatan * [1]

[1] Technische Universität Dresden = Dresden University of Technology – Germany

The communication of scientific uncertainty is a pressing issue in today's society, where public trust in science is crucial. My contribution to this discussion centers on the need to acknowledge and communicate uncertainty about scientific evidence in a way that maintains trust. While some argue that communicating uncertainty might foster skepticism or reluctance to act on scientific advice, I argue that it is essential to communicate uncertainty transparently. The challenge, however, lies in doing so without eroding public confidence in science.
To clarify this issue, I distinguish between **objective uncertainty** and **subjective or felt uncertainty**. Objective uncertainty refers to the indeterminacy present in the available evidence, which can be measured and explained. This is a natural aspect of scientific progress and should be acknowledged. However, felt uncertainty is a psychological state-an emotional response people have when faced with incomplete or ambiguous information. I argue that scientists should focus on communicating objective uncertainty while addressing the public's emotional response, ensuring that uncertainty is not perceived as a lack of credibility.

One of the central problems of conveying uncertain scientific messages is the emotional side of communication. People's distrust of uncertain scientific information is often rooted in emotional responses, such as anxiety or doubt, which can cloud their understanding of the actual evidence. This is where a **metacognitive account** of uncertainty becomes relevant. Metacognition involves individuals' ability to reflect on their own thought processes. When people feel uncertain, they are not only reacting to the information itself but also to their own doubts about whether they understand the situation. This can intensify feelings of mistrust.

My metacognitive approach suggests that addressing this emotional response is key to improving how uncertainty is communicated. Scientists should clearly differentiate between what is known and what is not, emphasizing that uncertainty is a normal part of scientific exploration. However, this objective uncertainty should be framed in a way that reassures the public rather than exacerbates feelings of doubt or fear. The literature on metacognitive feelings suggests that these emotional responses are often based on **social and cognitive cues** like ease of processing or familiarity, which can be purposefully integrated in the communication of scientific evidence. By fostering a better understanding of the process of scientific inquiry-where uncertainty often indicates areas of ongoing investigation rather than failure-while also communicating scientific evidence in a way that reassures people, scientists can reduce emotional resistance and build trust.
In conclusion, the communication of scientific uncertainty needs to balance the delivery of objective uncertainty with sensitivity to the emotional reactions it may provoke. My approach, which

*Speaker

highlights the metacognitive dimension of felt uncertainty, provides a framework for improving public engagement with uncertain scientific information. By openly addressing the uncertainty while reassuring the public, we can promote a more trusting and informed relationship between scientists and society.

Keywords: Science Communication, Uncertainty, Trust, Feeling of Uncertainty, Metacognition.

Statistical Learning Theory meets Formal Learning Theory - Ockham's Razor and VC Dimension

Joachim Stein * [1,2]

[1] Heidelberger Akademie der Wissenschaften – Germany
[2] Universität Tübingen – Germany

Ockham's razor is the principle that, all things being equal, one should favor simpler theories. The principle is fundamental to scientific hypotheses choice, has been studied in formal learning theory, and its role in machine learning has recently sparked much debate. In this article, I integrate the central hypothesis choice principle from statistical learning theory into the formal learning theory framework to analyze its connection to Ockham's razor.
The literature on formal learning theory provides a simplicity relation and defines an Ockham method as a method that selects the simplest hypothesis according to this simplicity relation. Ockham's razor is then understood as an application of an Ockham method (Genin and Kelly 2019).

The fundamental theorem of statistical learning theory states that the smaller the Vapnik-Chervonenkis dimension (VC dimension) of a hypotheses class, the better the hypotheses class can be learned. The central point of debate in the literature on the connection between statistical learning theory and Ockham's razor is whether the minimization of the VC dimension of a hypotheses class captures intuitive ideas about Ockham's razor, e.g. the Popperian idea that Ockham's razor selects the most falsifiable hypothesis (Corfield et al. 2009, Sterkenburg 2023)

In the paper, I investigate whether a VC dimension minimizing method is an implementation of Ockham's razor. Based on ideas in Steel (2009), I introduce a formal learning theory version of VC dimension and define a VC lowest method as a method that selects its output from the hypotheses class with lowest VC dimension compatible with the data. It turns out that VC lowest methods are not equivalent with Ockham methods. I show that in a special class of inductive problems, VC lowest methods satisfy more properties associated with the principle of Ockham's razor than Ockham methods, in particular the Popperian intuition that Ockham's razor selects the most falsifiable hypothesis. However, the results crucially depend on two assumptions: first that a strengthening of one's opinion counts as a mind-change and secondly, that the possibility to suspend judgment is not a necessary condition for a formalization of Ockham's razor. While Genin and Kelly (2019) strongly dismiss these assumptions, I discuss whether there are theory choice problems in which they are reasonable. I conclude that for a special class of inductive problems, the notion of a VC lowest methods conforms to the principle of Ockham's razor better than the notion of an Ockham method.

*Speaker

Literature:

Corfield, D., Schölkopf, B. & Vapnik, V. (2009). Falsificationism and Statistical Learning Theory: Comparing the Popper and Vapnik-Chervonenkis Dimensions. *J Gen Philos Sci* **40**, 51–58.

Genin, K. and Kelly, K. (2019). Theory Choice, Theory Change, and Inductive Truth-Conduciveness. Stud Logica, 107:949–989.

Steel, D. (2009). Testability and Ockham's Razor: How Formal and Statistical Learning Theory Converge in the New Riddle of Induction. Journal of Philosophical Logic, 38(5):471-489.
Sterkenburg, T. F. (2023). Statistical Learning Theory and Occam's Razor: The Argument from Empirical Risk Minimization.

Keywords: Statistical Learning Theory, Formal Learning Theory, Ockham's razor

Values in machine learning: What follows from underdetermination?

Tom Sterkenburg * [1]

[1] LMU Münich – Germany

Machine learning is in many ways biased. Much contemporary work in computer science as well as philosophy is devoted to charting the various types and entry points of algorithmic bias in machine learning pipelines. Several authors have further made a connection to the debate in the philosophy of science about the role of non-epistemic value judgments in scientific inference. Some have adopted arguments against the value-free ideal of science to reason more fundamentally that machine learning algorithms *must* be value-laden.

There arguments rely on the inductive nature of scientific inference, and the fundamental problem of the underdetermination of inductive conclusions by the available data. These are general characteristics that are shared by learning algorithms. Thus it has been claimed that "(t)hese arguments result in the view that both scientific and algorithmic decision procedures are deeply value-laden" (Johnson, 2024, J. Moral Phil.).

Yet there is something unsatisfying about leaving it at the conclusion that machine learning algorithms must be the product of value-laden choices beginning to end. What seems in order is a more careful picture of how both epistemic and non-epistemic factors come together in the design of machine learning algorithms. One step towards such a picture is to show why a general argument from underdetermination does *not* already settle the matter that machine learning algorithms must be value-laden. That is what I aim to show in my talk.

I will start by rehearsing the general argument from underdetermination. Next, I will set up my critique by clarifying and delineating the relevant notion of "value-ladeness" and of "machine learning algorithm." I consider here not the trained machine learning model (the usual focus in discussions of algorithmic bias), but the vehicle for the actual inductive inference itself, namely the actual learning algorithm. In terminology by van Fraassen, we are here concerned with the rationality rather than the vindication of a learning procedure.

Finally, I show why the underdetermination argument does not suffice to establish the valueladeness of learning algorithms. Here I rely on a distinction between domain-general learning rules and the domain-specific inductive models and corresponding inductive biases they must be provided with. I argue that while underdetermination does imply the need for such inductive biases, these do not have to (or even typically) come in the form of general Kuhnian "epistemic virtues." Rather, in line with Norton's material theory of induction, they can take the form of "local," domain-specific assumptions, that at least *could* encode an epistemically motivated assessment of a learning problem at hand. Further, I argue that the context of the design of

*Speaker

general learning rules is really the quintessentially epistemic project of philosophers and theoreticians to delineate principles of rational learning. I illustrate these observations both in the abstract (by means of the framework of Bayesian inductive inference) and in the more concrete (by the specific example of a convolutional neural network for digit recognition).

Keywords: machine learning, values, inductive inference, inductive underdetermination

Particle physics mechanisms and their explanatory value

Michael Stoeltzner * [1]

[1] University of South Carolina (USC) – United States

Textbooks of present-day elementary particle physics list several mechanisms, among them the Higgs mechanism or the see-saw mechanisms. This does not imply that quantum field theory were interpretable in terms of mechanical models or that scattering processes led to a return of causal explanation. But neither is 'mechanism' simply a metaphor for theoretically motivated reshufflings in Lagrangians. The aim of my paper is to argue that some of these mechanisms are explanatory in the sense of the late Wesley Salmon's mechanistic philosophy, based on probabilistic processes, and the recent account of mechanical explanation in biology and cognitive science advocated by Machamer, Craver, Dardeen, Glennan and others, where mechanisms are understood as the stable interaction of parts of a system that produce an empirically detectable phenomenon.

Given the probabilistic nature of quantum field theory, particle physics mechanisms often come in the form of how-possible or how-plausible explanations rather than providing how-actually explanations for scattering phenomena in terms of a fundamental causal mechanism. The above-mentioned biological accounts, however, insist that mechanisms are fundamental in a biological perspective – even if not reductionist in the sense of what Nicholson has called 'mechanicism'. As Kuhlmann and Glennan have argued, this lack of fundamentality poses an obstacle to the integration of the new mechanistic explanation into quantum theory. Examples of mechanisms in quantum theory, they argue, are either systems explicitly involving two distinct levels, such as lasers, or interpretations of quantum mechanics that yield determinate measurement outcomes on the basis of more fundamental processes, such as decoherence.

It seems to me that the recent debates about the explanatory value of effective field theories and about 'effective realism' (by Porter Williams and others) allow us to neglect fundamentality as long as the separation of scales that lies at the heart of renormalizable quantum field theories, is respected. While I do not follow Adrian Wüthrich's claim that Feynman diagrams automatically contain the multilevel structure required for the new mechanical explanation because they provide rules for local processes at higher loop order in the overall scattering process, Feynman diagrams express a basic feature of most mechanisms in quantum field theory: the interaction between actual and virtual particles. Virtual particles cannot be produced at the available energy but contribute to the higher orders of the scattering cross section. Since they accordingly contribute to observable phenomena, virtual processes do not represent pseudo-processes in the sense of Salmon. This distinction appears to me more basic than the question of fundamentality because by their very definition pseudo-processes cannot be explanatory. Not all theoretical objects appearing in Feynman diagrams contribute to processes, if virtual ones. According to most physicists, Faddeev-Popov ghosts are merely mathematical artifacts and the BRST quantization is not considered a mechanism. Interpreting particle physics mechanisms along Salmon's

*Speaker

lines requires understanding the often non-fundamental conservation laws and exclusion roles of particles as conserved quantities. In this respect the new biological accounts are less demanding because they do not require a formal theory of the parts of the explanatory mechanism.

Keywords: Explanation, mechanisms, quantum field theory, Salmon's mechanical philosophy

Start Making Sense: Understanding Particle Physics with Deep Neural Networks and Explainable AI

Frauke Stoll * [1]

[1] Technische Universität Dortmund [Dortmund] – Germany

As Machine Learning, particulary Deep Neural Networks (DNNs), is taking over science by storm, disciplines like Particle Physics are experiencing transformative shifts. While DNNs deliver impressive results, their opaque black-box nature presents a challenge to scientific understanding. Many prominent applications, including unsupervised anomaly detection and signal/background classifiers, provide accurate predictions but obscure the underlying physics, prompting the worry that scientists have shifted their focus away from understanding to prediction.
A promising solution lies in a specific kind of Explainable AI (XAI) method, which I refer to as scientific XAI (sXAI). Unlike traditional XAI, which seeks to explain how a DNN works, sXAI focuses on what the network has learned, addressing the problem of so-called what-opacity. However, how scientific understanding emerges from these methods has not been spelled out so far. Since scientific understanding is often linked to explanations, it is crucial to show how sXAI can provide such explanations. Analysing a case study from particle physics, I will demonstrate how DNNs contributes to scientific understanding when paired with sXAI

I will argue that understanding unfolds through several steps involving the DNN, sXAI, and the scientist. First, the trained DNN makes accurate predictions. Then, sXAI methods translate these predictions into a human-readable form, revealing the key features the DNN used. This, however, only addresses a "what-question": What did the DNN learn to make its decision? For a full explanation, answering the "why-question", this information must be physically interpreted and connected to theories. Finally, the model's results must be externally validated. These steps will be discussed in my talk, showing how DNNs and sXAI together allow scientists to construct an explanation of the phenomenon by providing them with explanatory information – the DNN is used to generate key ingredients for building more comprehensive models.

This case study leads to several insights. First, while DNNs are powerful tools for discovery, they do not inherently provide scientific understanding. sXAI can bridge this gap by offering intelligible representations that scientists can connect to physical theories. Second, the process of understanding begins with answering "what-questions"-identifying patterns and features in the data. Contrary to some claims in the literature, sXAI acts more as a translator than a full explanation. It is the scientist who must link sXAI's findings to broader scientific knowledge to answer the "why-question" and generate an explanation.
Ultimately, the process yields a scientific explanation comparable to traditional forms, though

*Speaker

it follows a different path. The joint effort of ML and scientists provides explanatory insights that can be evaluated against the same standards as other scientific explanations. In this way, the approach aligns with many established accounts of scientific understanding.

Keywords: Machine learning, Deep Neural Networks, Particle Physics, Explanation, Scientific Understanding, XAI

Present Records of the Past Hypothesis

Athamos Stradis * [1]

[1] King's College London – United Kingdom

Our world exhibits a 'record asymmetry' in the sense that it contains many records of the past but not of the future. However, its fundamental dynamical laws are time-symmetric. In these circumstances, how might the record asymmetry be explained? If time-asymmetry doesn't enter the picture via the dynamics, then it must do so via some sort of boundary condition. The standard candidate is the 'Past Hypothesis', the posit that the universe began in some particular very-low-entropy macrostate. But it's one thing to acknowledge this bare logic, and quite another to actually join the dots between the universe's initial state and the record asymmetry.

To this end, David Albert and Barry Loewer argue that because the initial macrostate occupies a specific tiny corner of phase space, it imposes a tight probabilistic constraint on the universe's possible past macrohistories, but not on its future macrohistories. By utilising records in the present, we exploit this constraint to infer the past, for this utilisation is steered by an evolutionarily hardwired, indirect acquaintance with the initial macrostate. Hence, the Past Hypothesis explains the constraint on past macrohistories, and this in turn explains the record asymmetry.

Although this picture has faced numerous objections, many misinterpret it as a literal account of human thought when dealing with records, and perhaps many more are aimed at the validity of the Past Hypothesis in the first place. In this paper I attempt to analyse Albert and Loewer's account in a way that's both charitable towards its content and sympathetic towards their background framework. Nevertheless, whilst I agree that there exists a probabilistic constraint on past macrohistories, I do not think they can make sense of it without helping themselves to the very records they wish to explain.

I will argue that this circularity is avoided if we explain the record asymmetry through a different route: via the 'fork asymmetry', a probabilistic structure connecting localised events. There are good reasons to think this phenomenon goes a long way to underwriting the effectiveness of records as windows on the past. This will warrant some brief discussion, but my main focus will be on why it's a better way of linking the record asymmetry to the universe's initial state. Neither Albert nor Loewer imagines their theory of records to be a complete one, and Albert in particular suspects that the fork asymmetry plays some role. But as we shall see, my way of looking at this structure leads to a very different understanding of the record asymmetry.

Keywords: Mentaculus, Past Hypothesis, Record Asymmetry, Time, Asymmetry, Statistical Me-

*Speaker

chanics

Kant's empirical account of science

Thomas Sturm * [1], Larissa Wallner * † [2]

[1] Institució Catalana de Recerca i Estudis Avançats = Catalan Institution for Research and Advanced Studies – Spain
[2] Goethe-University Frankfurt – Germany

We reconstruct Kant's empirical account of science, focusing on psychological, social, and historical aspects of scientific inquiry and progress. In section 1, we start by claiming that Kant's account presents a complex and dynamic image of the research process. Our reconstruction offers a more realistic portrayal of Kant's philosophy of science, revealing how these empirical conditions influence scientific progress and complement his critical philosophy. Section 2 looks at the psychological conditions, covering especially (1) the foundations of Kant's faculty analysis, (2) the distinction between higher and lower faculties, (3) the enhancement of observation, (4) the roles of imagination, (5) intellectual or reflexive aspects of scientific inquiry and (6) the special innate talent for scientific discovery, or "sagacity". Section 3, again, addresses important social and historical aspects of research, focusing on three different forms of the division of cognitive labor, and on Kant's approach to the history of science: (1) the general challenges science has to face (2) the division of labor between different kinds of minds (3) the societal prerequisites for flourishing scientific work and finally, (4) the importance of testimony for scientific progress. In conclusion, we reflect on how the empirical account relates to Kant's purely philosophical claims about science, and his view about how the history of science ought to be written.

Keywords: Psychology of science, sociology of science, history of science, scientific inquiry, scientific progress, Kant.

*Speaker
†Corresponding author: larissa.wallner@gmx.de

Cambridge Changes and Developing Immunity in the SARS-CoV-2 Pandemic

David Stöllger * [1]

[1] Universität Bielefeld – Germany

In this paper I examine the role 'Cambridge changes' play in disease classification. Following Guerrero (2010), I argue that classifications that rely on statistical normalcy may license 'Cambridge changes': a person's condition may undergo a shift in classification without the person having undergone any change themselves but is due to extrinsic factors.
This undermines Boorse (2014)'s claim that the prominent, naturalist Biostatistical Theory (BST) can solely rely on value-neutral, intrinsic pathology, i.e., below-normal functioning within a reference class. Specifically, it would not be a person's own deviation from normal functioning in comparison to a reference class, but rather an extrinsic change in other members that makes a difference. (Guerrero, 2010, p. 275) Boorse responds that if such changes occur, they are of no importance to medical practice. (Boorse, 2014, p. 35) Indeed, the highly abstract examples utilized by Guerrero (2010, p. 277) have little practical implications. Further, Cambridge changes need not imply that BST's notion of 'below normal function' is tied to external factors, but rather "(w)hat has changed is only species-typical functional capacity – the benchmark for whether the individual capacity is healthy or not." (Boorse, 2014, p. 36)

Still, I propose that Cambridge changes potential to shift this benchmark thus can make classifications less temporally stable and more susceptible to value influences than Boorse is ready to admit. In this paper I argue that our fast-adapting immune responses to infectious diseases provide a concrete example of a Cambridge change on a time scale relevant for classification:

In the beginnings of the recent pandemic susceptibility to symptomatic disease caused by SARS-CoV-2 was statistically normal in specific reference classes (e.g., older males, co-morbidities). (Cevik, Kuppalli, Kindrachuk, & Peiris, 2020, p. 4) Before late 2019, this susceptibility was not a diminished function impacting individual survival and reproduction, and thus, according to BST, was not a pathological condition. During the pandemic, however, susceptibility to symptomatic outcomes decreased, as many developed immunity (due to infections, vaccinations, pathogen mutation) or by succumbing to COVID-19.

Now, in the endemic phase susceptibility to symptomatic infection has become increasingly rare. (Bobrovitz, et al., 2023, p. 556)A successful immune response is now normal functioning; a failure of it is both a diminished and statistically abnormal functioning affecting individual survival and reproduction as expected in a reference class. Thus, at this point, a person's continued inability to mount an effective immune response is, according to BST, a pathological condition. But such a susceptible individual has undergone no pathological change, rather there has been both a change in external environment (i.e., infectious entity prevalence and mutation)

*Speaker

and a change in degrees of immunity in the population and each reference class.
This concrete example is on a sub-evolutionary timescale and may lend genuine support to Guerrero's thesis. Immune-competent persons developed functional capacities that potentially affect the designation of other persons, who have undergone no change in functional capacities, as 'immuno-compromised'. Interestingly and as indicated above, vaccines may play a role in such a Cambridge change.

Keywords: Philosophy of infectious disease epidemiology, philosophy of medicine, concept of disease, Cambridge Change

How Many Priors and Inductive Biases Can One Afford in the Neo-Empiricist Dogma?

Vanja Subotić * [1]

[1] Institute of Philosophy at the Faculty of Philosophy, University of Belgrade – Serbia

In cognitive science and AI research, the perennial rationalism vs. empiricism debate has taken the form of domain-general vs. domain-specific mechanisms thereby influencing the methodological decisions of using either symbolic or connectionist models to explain human cognition. As Margolis & Laurence (2013, 2015) point out, the two camps differ with respect to styles of learning-based explanations: empiricist insists on a small number of mechanisms that are generally and repeatedly engaged to process different inputs, whereas rationalists think that a plethora of pre-wired specialized modules for different inputs guides the development of our cognitive capabilities. In my talk, I discuss how many priors and inductive biases one can allow in the connectionist computational architecture to defend the empiricist account of cognitive architecture consistently (cf. neo-empiricist DoGMA in Buckner 2024).
Marcus (2018), a long supporter of nativist ideas provides a list of computational primitives that any computational architecture that aims at faithful representation of human cognition must incorporate: **(a)** Representations of objects, **(b)** Structured, algebraic representations, **(c)** Operations over variables, **(d)** A type-token distinction, **(e)** A capacity to represent sets, locations, paths, trajectories, obstacles, individuals, **(f)** A manner of representing the affordances of objects, **(g)** Spatiotemporal contiguity, **(h)** Causality, **(i)** Translational invariance, **(j)** A capacity for cost-benefit analyses, **(k)** Representation of time, **(l)** Intentionality.

Marcus does not distinguish domain-specific mechanisms or capacities from domain-general ones, nor priors from inductive biases, but rather gives the impression that **(a)**-**(l)** must be innate or connectionist models are good for nothing. However, **(a)**, **(e)**, **(f)**, **(g)**, **(h)**, and **(k)** are linked to core cognition, or, more precisely, intuitive physics developed from infancy, and can be considered domain-general, although the format of (a) is for Marcus probably symbolic rather than distributed as in connectionism. On the other hand, **(b)** and **(c)** amount to a strong theoretical commitment that cannot be incorporated into connectionist models if the banners of moderate empiricism are to be defended given that such innate mechanisms are conceived as domain-specific. Moreover, **(b)** and **(c)** cannot be incorporated into any connectionist model if connectionism strives to be accepted as an autonomous research program of human cognition. Thus, Marcus demands the impossible.
The issue at hand is whether the empiricist can show that intuitive physics can be incorporated into a connectionist model, albeit with fewer priors than Marcus would presuppose. Piloto et al. (2022) have presented a novel connectionist model based on a deep learning algorithm that yields the mechanisms for intuitive physics concept acquisition underlying infant visual cognition. The team combined developmental psychology with engineering and through the violation-of-expectation paradigm. As long as the model remains object-centered, it can de-

*Speaker

velop robust violation-of-expectation effects with a strikingly small amount of training data-*cca* 28h of visual experience. Piloto et al. explicitly state that their model is successful in virtue of having two priors (along with inductive biases such as convolution, distributed representations, and DL algorithm, cf. Goyal & Bengio 2020: 5), albeit not in the form of object representations as Marcus would prefer.

Keywords: connectionism, deep learning, empiricism, inductive bias, intuitive physics, prior.

Boundary Conditions and Conditions at the Boundary

Jeffrey Sykora * [1]

[1] CHI Saint Joseph Health – United States

Boundary conditions and conditions at the boundary are not the same thing. The distinction between these two concepts, described by Brenner and Ganesan (2000) in the context of modelling fluid systems, has important implications for philosophical accounts of scientific explanation and intertheory relations. The distinction, put briefly, is that conditions at the boundary, but not boundary conditions, explicitly describe interactions between a fluid and a boundary. While boundary conditions define a constraint on the behavior of a fluid at a boundary, conditions at the boundary arise from the interactions between the fluid and the boundary. The distinction also has practical importance for the construction of scientific models. Indeed, Brenner and Ganesan's case for the distinction in the first place shows how conditions at the boundary described by a molecular model do not always correctly predict the boundary conditions of a corresponding continuum model.
In this context, this distinction might initially seem like a mere expression of an obvious difference between fluid dynamics and molecular dynamics. That is, it might simply appear that fluid dynamics uses boundary conditions in its models, while molecular dynamics uses conditions at the boundary. However, I show that the respective concepts are not tied necessarily to a particular model or even to a particular scale. In fact, on the one hand, boundary conditions are used in molecular models, as with the use of a tangential momentum accommodation coefficient to define the behavior of fluid molecules at a solid wall. And on the other hand, conditions at the boundary are used in continuum fluid models, as with descriptions of wetting behavior of liquids on solid surfaces, which depend on the interaction between the physicochemical properties of both the liquid and the solid.

Boundary conditions and conditions at the boundary play different respective roles in scientific explanation. This is because of the particular relations in which they stand to the other parts of the models they help define. On the one hand, conditions at the boundary provide information about the boundary itself that boundary conditions do not, which allows them to play some explanatory roles that boundary conditions cannot. On the other hand, boundary conditions have a lawlike character, which conditions at the boundary lack, that allows them to stand in particular kinds of explanatory relations to the governing equations of fluid dynamics.

The distinction between boundary conditions and conditions at the boundary is also relevant to intertheory relations. It might at first seem to support reductionist views of intertheory relations. However, the provided examples show that the decision of whether to use boundary conditions or conditions at the boundary does not essentially depend on the scale of a system. Even though, in some sense, fluid dynamics depends on the underlying molecular nature of flu-

*Speaker

ids, we cannot assume that those molecular facts alone generally determine the fluid dynamic boundary conditions.
While the present discussion depends on examples from fluid systems, its conclusions might inform analysis and practice in other contexts in which boundaries play a role.

Keywords: boundary conditions, fluid dynamics

Evaluating evidence of mechanism in psychiatry: implications of a failed extrapolation from biomedicine

Yiran Tao * [1]

[1] Department of Science and Technology Studies, UCL [London] – United Kingdom

The Russo-Williamson thesis (2007) states that we need both mechanistic and difference-making evidence to establish causal claims in biomedicine. Illari (2011) refines this by differentiating between *evidence acquired through* a mechanistic or difference-making approach and *evidence of* mechanisms or difference-making. Illari proposes a framework for evaluating *evidence of* mechanisms and its strength in complementing *evidence of* difference-making to support causal claims. I will argue that this framework has its limitations in dealing with complex medical conditions where there are conflicting bodies of evidence. I will give a preliminary analysis of a case in psychopharmacology, the selective serotonin reuptake inhibitors (SSRIs). I will lay out three reasons why the framework falls short in this case and put forward some implications for scientific pluralism regarding special sciences.
The serotonin hypothesis of depression postulates that depression is caused, at least partly, by abnormalities in the brain's serotonin system. SSRIs are serotonergic antidepressants designed to block a neurone's re-absorption of serotonin, thus increasing the concentration of serotonin in the brain. Recently, however, an umbrella review (Moncrieff et al. 2022), bringing together existing systematic reviews and large database research on the serotonin hypothesis, concluded that no consistent neurobiological evidence supports the serotonin hypothesis.

The review primarily analysed correlational and experimental studies lacking robust difference-making evidence for serotonin's role in depression. Other relevant bodies of evidence that were not presented in the review include difference-making evidence of SSRIs and mechanistic evidence of serotonin and SSRIs. Together, these bodies of evidence cover most categories of evidence Illari considered in her framework. If we carefully integrate the evidence, we should be able to confidently state whether serotonin causally contributes to depression or not. However, such integration seems beyond attainment. The key contention is whether different bodies of evidence can be used to corroborate or undermine each other. Sceptics doubt the presumptions that different bodies of evidence are evidence of the same stand-alone process that has a linking mechanism capable of difference-making (Kendrick & Collinson 2022; Science Media Centre 2022).
I argue that such scepticism is warranted, and evidence of mechanism has very limited power in corroborating the evidence of difference-making in this case. This contrasts with the conclusion we may draw if we evaluate these pieces of evidence with Illari's framework. I give three reasons for this. First, we do have evidence of mechanisms around serotonin. However, whether they are part of the linking mechanism is unknown. Second, the postulated linking mechanism between

*Speaker

serotonin and depression involves ill-defined entities and activities in psychiatry. Third, it is difficult to characterise depression as a collection of entities and activities to allow it to be the end condition of a mechanism. I will consider some counterarguments that allow a modified Illari's framework to work for psychiatry and psychopharmacology. A further implication of this analysis is our two options towards special sciences like psychiatry, a unificationist one and a pluralistic one. Some counterarguments have unificationist connotations, while I propose a pluralistic stance (Kellert, Longino, & Waters 2006).

Keywords: Evidence, Mechanism, Psychiatry, Causation, Pluralism

Metaphysical Trouble for Logical Abductivism

Iulian Toader * [1]

[1] University of Vienna [Vienna] – Austria

Logical abductivism maintains that the rational choice of a logical theory follows a methodology which is similar to that employed in choosing between scientific theories: one selects the theory with the highest rationality index, which reflects the theory's performance on criteria such as adequacy to the relevant data, simplicity, etc. The empirical sciences have been typically taken to provide abductive evidence for classical logic (CL). Nonetheless, standard quantum mechanics (QM) has been sometimes understood to provide evidence for some non-classical logics like non-distributive quantum logic (QL), and this can be naturally considered as abductive evidence for QL.

This conflict appears to have been solved by pointing out that CL is rationally preferable to QL because it is more adequate to the relevant data, and in particular to the metaphysical consequences of QM. In my paper, I will discuss some reasons for doubting this adequacy claim. My main goal is not to defend QL as a logical theory that is rationally preferable to CL; rather, I want to make the point that the logical abductivist must have a hard time choosing.

Note that the choice of a logical theory can proceed abductively only if one assumes the same conception of logic and thus, arguably, only if one agrees on the relevant data for abduction. It is plausible to believe that QL could be preferred to CL, if the CL and QL advocates were not to share the same conception of logic. However, even if one assumes the same conception of logic and thus, arguably, agrees on the relevant data, abduction may fail to guide one's rational choice. This can be illustrated by assuming an inferentialist conception of logic, according to which the meanings of logical terms are determined by their rules of inference. But perhaps this only indicates that the abductivist should not be an inferentialist.

More importantly, I will argue that even if one assumes the same conception of logic, one may nevertheless disagree on the relevant data, and thus abduction would still be unable to guide one's rational choice. This will be illustrated by assuming a metaphysical conception of logic, according to which logical consequence is a nonmetalinguistic relation between worldy facts. In this case, the abductivist's choice between CL and QL must remain elusive for the following reason: if actual metaphysical consequences of QM are the only relevant data, CL is rationally preferable to QL; but if non-actual possible metaphysical consequences of QM are also considered as relevant data, then QL is rationally preferable to CL.

One cogent objection that I will have time to address goes as follows: if both actual and non-actual possible metaphysical consequences of QM are the relevant data, then their abductive

*Speaker

explanation offered by the QL advocate is nonsensical. For what might it mean to say that the distribution laws are correct, but it is possible that they are incorrect? I will show how the abductivist can consistently make sense of this explanation as an epistemic contradiction.

Keywords: logical theory choice, classical logic, quantum logic, quantum metaphysics, epistemic contradictions

The explanatory power of network dynamics

Dingmar Van Eck * [1]

[1] University of Amsterdam [Amsterdam] = Universiteit van Amsterdam – Netherlands

A widely endorsed view in the philosophical literature on scientific explanation subscribes to the idea that the explanatory power of scientific explanations resides in their ability to track counterfactual dependencies (e.g., Woodward 2003; Saatsi and Pexton 2013; Rice 2015; van Eck and Wright 2021; Kostic and Khalifa 2023). What also appears to be widely endorsed is the notion that models used for explanatory purposes should articulate and enable tracking (only) *robust* dependency relations (e.g., Woodward 2003). The robustness of the modeled dependency relations is taken to be a proxy for the explanatory power of a model-based explanation.
In this contribution I argue that a number of scientific modeling cases do not fit this view. In these cases*, things are exactly the opposite*: *the more a model enables adapting and shaping explanations across different contexts (by highlighting changing dependency relations), the higher its explanatory power*.

I build my analysis by relating it to examples drawn from modeling in developmental psychology (Thelen et al. 2001), psychopathology (Borsboom 2017) and molecular neuroscience (Bollhagen and Bechtel 2022). Consider e.g., network models in psychopathology. The network theory of mental disorders conceptualizes mental disorders in terms of networks of causally connected symptoms. Mental disorders are understood as dysfunctional states in which such networks can get locked. One major aim of network theory is to (statistically) estimate such networks on the basis of data and build *computational* network models of mental disorders (Borsboom 2017).

A key insight is that mental disorders/network structures change and evolve across time and across agents. Hence, *counterfactual dependencies between symptoms are also dynamical and prone to change*. The point of network modeling is to identify the relevant symptoms, and to clarify how they may change, as conditions change, or to clarify how these symptoms and their relations differ across cases (Robinaugh et al., 2019). For instance, in the case of panic attack and disorder, clarifying how the impact of arousal on perceived threat is different across cases in light of different beliefs about the threat of a specific stressor (e.g., a growling dog) and/or different appraisals of arousal levels across cases. In this example, the symptoms composing the causal structures differ across cases, i.e., different beliefs and/or different appraisals. So solely focusing on (a model enabling the tracking of) robust dependencies as a proxy for explanatory power of the network model runs counter to explanatory purposes in the case of network modeling. If anything, the contrary is the case: *the more a network model is able to adapt and shape explanations across different dependency relations, the higher its explanatory power.* Tracking *changing* (context-sensitive) dependencies is also key.
I argue that this point generalizes and applies to other cases and types of explanation, to wit: specific dynamical models in developmental psychology (Thelen et al. 2001) and specific models

*Speaker

of molecular mechanisms in neuroscience (Bollhagen and Bechtel 2022). The bigger point is that philosophical theories of explanatory power should be extended to also cover these cases.

Keywords: scientific explanation, explanatory power, model

Against Grounding by Absence

Martin Voggenauer * [1], Kian Salimkhani [2]

[1] University of Cologne – Germany
[2] Radboud University Nijmegen – Netherlands

Grounding is often regarded as a form of metaphysical causation. However, while in the case of causation, relata are generally considered to be events, the question of what entities should be considered as relata of the grounding relation is more controversial. Thus, the question arises whether there are any constraints on what entities can feature as fundamentals or grounds. In particular, by analogy with the concept of negative causation, the question arises whether grounds can be negatives, or whether only positive or actual entities can serve as grounds.
Arguably, grounding by absence or negative grounding can be understood as grounding by negative facts. So, a reluctance to accept grounding by absence can be traced back to an attitude quite common in philosophy that "negation is at best a nifty shorthand, (...) not a way to get at fundamental reality" (Muñoz 2019, 1). According to this view, negative facts are non-fundamental, because they require "no extra work" (Muñoz 2019, 4). Thus, the question is whether grounding by absence is sensible or whether (physical) entities require positive building-type relations à la Bennett (2017) without reference to absences. Specifically, we can ask whether building-type relations that connect different levels of nature are always non-absence relations.

To answer this question, we turn to the related debate on causation by absence (or omission) and examine what we can learn from it for the case of grounding by absence. On the one hand, a possibility to defend grounding by absence might employ the analogy to negative causation as follows: Mumford & Anjum (2011, 56–58) argue that in theories of causation without absence conditions, causes can never be sufficient. Sufficient conditions must therefore always include the absence of preventing factors. Muñoz (2019) uses a similar argument for grounding. On the other hand, it is possible to argue against grounding by absence by referring to the problems that negative causation generally faces. In addition to the general question of how omissions can cause anything at all, these problems include the fact that negative causes are highly context-dependent. Accordingly, Beebee (2004) argues that omissions cannot be relata of causal relationships since they do not cause anything. Instead, she suggests that omissions can only figure in causal *explanations* by relying on normality conditions.

In this talk, we will argue against absences as relata of the actual grounding relationships to overcome the problem of "no extra work". Instead, in analogy to Beebee's argument on negative causation, we will suggest that absences can only play a role for explanatory purposes (i.e., in what we might call metaphysical explanations, as opposed to causal explanations).

References

Beebee, Helen (2004): Causing and Nothingness. In J. Collins, E. J. Hall & L. A. Paul (Eds.),

*Speaker

Causation and Counterfactuals, Cambridge, MA: MIT Press, 291-308.

Bennett, Karen (2017): *Making things up.* Oxford: Oxford University Press.

Mumford, Stephen & Rani Lill Anjum (2011): *Getting causes from powers.* Oxford: Oxford University Press.

Muñoz, Daniel (2020): Grounding nonexistence. *Inquiry*, 63, 209-229.

Keywords: grounding, causation, fundamentality, negative facts, omissions

A Morality Evolutionary Game Theory Can Model

Mikhail Volkov * [1]

[1] MCMP, LMU Munich – Germany

Evolutionary game-theoretic (EGT) models of morality have been the target of several distinct and powerful criticisms. These aimed to show the models' explanatory irrelevance and general inability to support the hypothesis about moral emergence as a rational collective response to certain evolutionary pressures. The project of using EGT for the vindication of this hypothesis has largely stalled in the last decade, with the impression being that the criticisms were ultimately successful in showing that EGT simulations are bound to remain toy models with little connection to reality.
The pessimistic impression is premature. The work submitted for the conference tries to argue so by being the first thorough attempt to defend EGT models against the criticisms and supplying a larger strategy regarding how these modes can be modified to achieve a more serious explanatory import. I defend the thesis that the criticisms ultimately show only that the existing models in the EGT tradition are what philosophers of models call epistemically opaque how-possibly models. This diagnosis, however, is by no means lethal to the general project of explaining the existence of morality using simulations. In particular, the latter can be modified to reach the status of causal how-possibly models, thus supplying real resources for philosophical arguments about moral emergence rather than empirically irrelevant formal exploration.

On the basis of modelling literature, I outline four steps towards how such an update can be achieved. Crucially, I present tangible suggestions towards satisfying three of these. These are mainly concerned with fostering closer collaboration between palaeoanthropology and evolutionary game theory and drawing on quantitative evidence from the former to improve models of the latter, as well as defining a clear sense of "morality" that computational models can handle. Concrete parameter ranges and new idealisations for model design are provided, and archaeological and palaeoanthropological justification for these is given.

The work contributes two general arguments. One: whereas EGT models of moral evolution are shown inadequate by existing criticisms, this does not showcase the futility of using EGT in philosophical work concerning moral evolution. Two: there are concrete, implementable ways to make these models better; the project from which the conference submission has grown provides both theoretical justification and practical examples of such improvements.
The ultimate aim of the project is to chart a research direction for future modelling efforts using EGT methods with an aim to demonstrate that morality is an evolutionary product that emerged due to conferring evolutionary advantage on human populations. Better, more informative models in the EGT tradition can yet be produced to strengthen this hypothesis.

*Speaker

Keywords: game theory, evolutionary game theory, models, simulations, computational modelling

Biases in Causal Inferences

Lilian Von Bressensdorf * [1]

[1] Ludwig Maximilian University [Munich] = Ludwig Maximilians Universität München – Germany

The reliability of causal inference in complex empirical scenarios is a pressing issue, especially in light of pragmatics. For example, a physician prescribes a drug to a patient, and the patient undergoes various changes in symptoms. Physicians make an inference that the drug caused those changes on which they base treatment of subsequent patients. Yet, assessment of causal inference here depends not only on the causal relation between the drug's physiological effectiveness and the patient's symptom changes. One factor that may mislead causal inference is alternative causes that might also impact a patient's health improvement, such as the placebo effect. Another concern is that doctors might hold cognitive biases, which might distort the evaluation of patients' responsiveness to drug treatment.
In recent work, Tabatabaei Ghomi and Stegenga (under review) investigate causal inference in clinical expertise by computational modelling: A simulated doctor is confronted with a particular disease case. The doctor has a set of several drugs to prescribe to simulated patients without knowing their physiological effectiveness. Yet, by observing patients' improvement after drug treatment, the doctor uses causal inference to asses drugs' physiological effectiveness. Nevertheless, in addition to drugs' physiological effectiveness, patients' improvement might also be caused by the placebo effect and the natural course of the disease. In this way, Tabatabaei Ghomi and Stegenga's (under review) model considers the first misleading factor of causal inference applied to clinical expertise.

Results suggest that in many situations, the disease's natural course and the placebo effect were misleading causal inferences. Doctors overestimate drugs' physiological effectiveness. However, Tabatabaei Ghomi and Stegenga's (under review) model disregards the second distorting factor in causal inference: Doctors' cognitive biases. In this contribution, we extend Ghomi and Stegenga's (under review) model by three instances of cognitive biases: The confirmation bias, the background modulation bias, and the anchoring bias. *Confirmation bias* means doctors overestimate any treatment outcomes that support their expectations. *Background modulation bias* means that doctors modulate any treatment outcome in light of background beliefs about placebo effect and the natural course. And, under *anchoring bias,* doctors' hold prejudices on drug effectiveness before the first treatment. We evaluate the reliability of biased causal inference for eight different disease scenarios – each characterised by a different combination of drugs' physiological effectiveness, the placebo effect and the natural course of the disease.
From simulations we obtain, unsurprisingly, that confirmation bias reinforces the overestimation of drugs' effectiveness of unbiased causal inference. Background modulation bias leads to underestimation. Hence, in both forms, biased causal inference is rather unreliable applied to clinical expertise.However, the anchoring bias, a combination of confirmation bias and background modulation bias, and the combination of all three biases increase reliability of unbiaed causal inference. Here, cognitive biases mitigate the misleading impact of the placebo effect and

*Speaker

disease's natural course in causal inference. Thus, compared to the earlier model that means biased doctors may outperform unbiased doctor, which gives causal inference partial vindication in a clinical setting.

Keywords: Biases, Causal Inference, ABM

Counterfactual Pandemic Scenarios: An interventionist approach

Marcel Weber*†[1] and Raphael Scholl‡[1]

[1]University of Geneva – Switzerland

Abstract

Scientific policy advice often relies on scenario analysis, which uses computational models to provide projections about what *would* happen if specific policy options *were* adapted (or not). As a salient example, consider climate scientists' projections about how various climatic variables would be affected by different future levels of carbon dioxide emissions. Or consider the case on which we focus here: Epidemiologists' projections about how different public health measures – such as lockdowns, school closures, or vaccination campaigns – would affect Covid-19 case numbers. Such projections are typically counterfactual in that they describe possibilities that have not been and in many cases never will be realized.
How should we think about the nature and the empirical warrant of such counterfactual projections? The philosophical literature offers us several approaches for thinking about counterfactual assertions in, and counterfactual results derived from, scientific models. These include fictionalism, possibilism, and Lewis-Stalnaker counterfactual semantics. Here we present a novel approach that draws on Woodward's interventionist account of causation. Woodward has argued that ordinary causal claims are best interpreted as already having counterfactual content. To say that A is causally relevant to B is not simply to say that in an experiment, B appeared when A was realized in the intervention arm but did not appear when A was suppressed in the control arm. Rather, the causal claim includes the counterfactual content that if A *had* been realized in the control, then B *would* also have appeared in the control. The purpose of well-designed experiments (controlled, non-fat-handed, etc.) is to provide evidence for the truth or falsity of precisely such "interventionist counterfactuals".

On the basis of this account of interventionist counterfactuals, we argue for two related theses. First, the major claims of scenario modeling in infectious disease epidemiology (such as "school closures would reduce influenza cases by 10%") can themselves be understood as interventionist counterfactuals. However, their truth or falsity is difficult to assess by direct experimental intervention on suitable populations. Second, the computational models that are used for scenario modeling aim to find the truth values of population-level interventionist counterfactuals indirectly: by simulating the cumulative effect of causal processes at epistemically more easily accessible lower levels (e.g., by modeling person-to-person transmission and the movement patterns of the population between different locations).
Without defending any particular infectious disease model or modeling approach, we thus show that the counterfactual projections of scenario modeling are continuous with the routine scientific practice of causal inquiry. While the relevant disease transmission models may fail to represent their target systems adequately (as we will discuss), their counterfactual content

*Speaker
†Corresponding author: marcel.weber@unige.ch
‡Corresponding author: raphael.scholl@unige.ch

is non-mysterious in that it derives directly from the counterfactual content of the causal claims about disease transmission from which such models are built. We will argue that our interventionist counterfactual account can be seen either as a much-needed complement to other accounts of counterfactual modeling, or as a metaphysically leaner and practically more relevant alternative.

Keywords: scenarios, modeling, prediction, counterfactuals, interventionism, covid 19 pandemic

Counterfactual Pandemic Scenarios: An interventionist approach

Marcel Weber *† [1], Raphael Scholl * ‡ [1]

[1] University of Geneva – Switzerland

Scientific policy advice often relies on scenario analysis, which uses computational models to provide projections about what *would* happen if specific policy options *were* adapted (or not). As a salient example, consider climate scientists' projections about how various climatic variables would be affected by different future levels of carbon dioxide emissions. Or consider the case on which we focus here: Epidemiologists' projections about how different public health measures – such as lockdowns, school closures, or vaccination campaigns – would affect Covid-19 case numbers. Such projections are typically counterfactual in that they describe possibilities that have not been and in many cases never will be realized.
How should we think about the nature and the empirical warrant of such counterfactual projections? The philosophical literature offers us several approaches for thinking about counterfactual assertions in, and counterfactual results derived from, scientific models. These include fictionalism, possibilism, and Lewis-Stalnaker counterfactual semantics. Here we present a novel approach that draws on Woodward's interventionist account of causation. Woodward has argued that ordinary causal claims are best interpreted as already having counterfactual content. To say that A is causally relevant to B is not simply to say that in an experiment, B appeared when A was realized in the intervention arm but did not appear when A was suppressed in the control arm. Rather, the causal claim includes the counterfactual content that if A *had* been realized in the control, then B *would* also have appeared in the control. The purpose of well-designed experiments (controlled, non-fat-handed, etc.) is to provide evidence for the truth or falsity of precisely such "interventionist counterfactuals".

On the basis of this account of interventionist counterfactuals, we argue for two related theses. First, the major claims of scenario modeling in infectious disease epidemiology (such as "school closures would reduce influenza cases by 10%") can themselves be understood as interventionist counterfactuals. However, their truth or falsity is difficult to assess by direct experimental intervention on suitable populations. Second, the computational models that are used for scenario modeling aim to find the truth values of population-level interventionist counterfactuals indirectly: by simulating the cumulative effect of causal processes at epistemically more easily accessible lower levels (e.g., by modeling person-to-person transmission and the movement patterns of the population between different locations).
Without defending any particular infectious disease model or modeling approach, we thus show that the counterfactual projections of scenario modeling are continuous with the routine scientific practice of causal inquiry. While the relevant disease transmission models may fail to

*Speaker
†Corresponding author: marcel.weber@unige.ch
‡Corresponding author: raphael.scholl@unige.ch

represent their target systems adequately (as we will discuss), their counterfactual content is non-mysterious in that it derives directly from the counterfactual content of the causal claims about disease transmission from which such models are built. We will argue that our interventionist counterfactual account can be seen either as a much-needed complement to other accounts of counterfactual modeling, or as a metaphysically leaner and practically more relevant alternative.

Keywords: scenarios, modeling, prediction, counterfactuals, interventionism, covid 19 pandemic

Representing and Understanding with Feynman Diagrams

Karla Weingarten * [1]

[1] Radboud University Nijmegen – Netherlands

Feynman diagrams have come a long way since their inception in the 1940's. Originally constructed to help the calculation of quantum electrodynamic scattering terms, they provide a quick and straightforward way to derive individual terms of a perturbation series that approximates a scattering amplitude for a particle interaction.
Since then, Feynman diagrams seem to have developed a life of their own, their use being taught independently from the calculation of scattering amplitudes and their publication as figures independently of the derivation of perturbation terms. Physicists refer to them when discussing particle processes without referencing any calculations. An apparent development is visible: from mere calculation tools for complex approximation schemes to stand-alone representations of particle phenomena.

This development has not gone unnoticed by philosophers of physics. The received view argues that because Feynman diagrams are a calculational device, they can never stand for actual particle processes. Each diagram corresponds to only one of infinitely many contributions to the scattering amplitude. Even more, the diagrams depict so-called virtual particles, which are often considered as mediators of the force involved, but which cannot be localized experimentally. There is thus no basis on which Feynman diagrams, especially individual ones, can represent a particle interaction. The majority of philosophers are thus of the opinion that the majority of physicists are misguided by their own achievements.

In this talk, I argue that physicists do not use the diagrams in the way philosophy accuses them of. The problem that philosophers highlight stems from the diagrams' intricate connection to perturbation theory. Physicists, when using Feynman diagrams in a detached manner, are however not concerned with perturbation theory. They do not refer to the diagrams' power as calculation tools, but as something that provides epistemic access to the particle interaction. To address this disparity and account for physics practice, I conceptualize the diagrams as models representative of particle interactions. As I will conclude that no account of scientific representation available in the literature is suitable for Feynman diagrams, an alternative will be constructed, centered around the epistemic aim that is to be achieved with the model: A model M accurately represents a target T iff M provides understanding of T to a competent agent. Considering Feynman diagrams as models, we can employ a non-literal interpretation with which we can grasp them as representations of particle phenomena if they provide understanding of them. Importantly, this interpretation allows the explicit depiction of virtual particles as productive idealizations of the particle process because it is exactly them that provide epistemic access to that process. The explicit depiction of virtual particles and the consequences that this style

*Speaker

of notation has provide an understanding of particle processes unmatched by other representative vehicles. Their ability to facilitate counterfactual reasoning allows Feynman diagrams to provide modal understanding of their associated target. It is because of this ability to facilitate understanding that I conclude Feynman diagrams are representative of particle processes.

Keywords: Feynman diagrams, Models, Representation, Understanding, QFT

Weak Supplementation in Aristotle and Euclid

Benjamin Wilck * [1]

[1] PhD student, Humboldt University Berlin – Germany

In this paper, I mobilize the Greek arithmetical notion of a unit (μ) as found in Euclid's *Elements* to defend the mereological notion of parthood in Greek maths against some recent commentators who claim that it violates the Principle of Weak Supplementation.
Weak Supplementation simply means that if x is a part of y, some part of y will be left over. That is, whenever a whole has a part, it has another part that does not overlap with, and is hence mereologically disjoint from, the first (Simons 1987, 28; Cotnoir & Varzi 2021, 46; cf. Cotnoir 2021). Simons (1987, 116), Varzi (2008, 110), Koslicki (2008, 168), Corkum (2015, 804), among many others, hold that Weak Supplementation is analytic and must hence be satisfied by any genuine mereology.

Vlasits (2017, 138–139; 2019, 10–11) argues that Weak Supplementation is violated by the notion of parthood in Greek maths because there is no proper part of, say, 3 that does not overlap with 2. Indeed, according to both Aristotle and Euclid, 2 is a non-aliquot part, but still a proper part, of 3. Yet, since both Aristotle and Euclid deny that 1 is a number, there is no other number in 3 besides 2. If Vlasits is right, then the mereological theories of both Aristotle and Euclid are seriously flawed.

Against this, I maintain that there is a proper part of 3 that does not overlap with 2, – namely, the unit.

Yet one might want to object – as do Høyrup (2004, 143–144) and Borzacchini (2007, 289) – that Euclid excludes the unit from being a proper part of any number, given that his definitions of aliquot and non-aliquot parthood of a number (*Elements* VII, Def. 3) are restricted to numbers, whereas units are no numbers (*Elements* VII, Defs. 1–2; cf. Aristotle, *Metaphysics* N.1, 1088a5–7).

I suggest that this objection can be met by showing that a unit is a proper part of any number. This is implied by the definition of the number as "that multitude which is composed of units" (*Elements* VII, Def. 2), as well as by the fact that the unit measures any number (*Elements* VIII.6). The unit is in fact as much an aliquot part of 3 as, say, 2 is an aliquot part of 4.
After blocking other likely objections to the view that the unit is a proper part of any number, I conclude that the notion of parthood in Greek maths is consistent with Weak Supplementation.

*Speaker

Keywords: Euclid, Aristotle, Ancient Greek philosophy, Ancient Greek mathematics, Number Theory, Units, Mereology, Weak Supplementation

"We will read your emails": Navigating privacy concerns in research on modern scientific collaborations

Rebecka Mähring [1], Adrian Wüthrich *† [2]

[1] University of Cambridge – United Kingdom
[2] Technische Universität Berlin – Germany

Digital communications—conducted over email lists, chat forums, wiki pages, and more—are ubiquitous in modern scientific collaborations. Furthermore, digital communications constitute a unique component of how collaborations function which cannot easily be supplanted or approximated by other data sources such as, e.g., on-site participant observation. Therefore, these "born-digital" sources are an essential component of studies of modern scientific collaborations. Accordingly, researchers in the history and philosophy of science (HPS) must learn how to work with this type of data. On one hand, this implies that HPS researchers should upgrade their methodological toolbox with tools from the computational humanities which allow the analysis of large amounts of data. On another hand—and this will be the focus of our contribution to this conference—the HPS researchers will have to learn how to navigate privacy concerns and data protection regulations to conduct their study in an ethically (and legally) sound way.

For participant observation and similar methods, the "principle of informed consent"—that all potential participants must be fully informed of the study and thereafter asked individually if they consent to participate—is often considered a cornerstone of research ethics. However, in the case of large scientific collaborations, obtaining such consent is practically impossible for several reasons. One obstacle is the sheer size of many modern scientific collaborations. In addition, the membership of many collaborations fluctuates, which means that individual consent must be sought continuously from a large number of people. Furthermore, sampling a manageable subset of the collaboration for seeking consent is problematic, since this approach requires prior knowledge of or assumptions regarding the relevant roles and subgroups within the collaboration.

Therefore, it is impractical and inadequate to use informed consent to motivate the study of the internal communications in modern scientific collaborations. Other legal and ethical bases are needed. In our contribution, we argue that the General Data Protection Regulation (GDPR) of the European Union is sufficiently flexible to accomodate HPS research on the digital communication within collaborations. Moreover, we find that the approach which results from exploring the various legal bases for the necessary processing of personal data is, in many ways, a continuation of best practices which have been discussed and implemented by researchers before the advent of GDPR. Accordingly, the GDPR turns out to be a useful framework, which can help researchers conduct new types of HPS studies on scientific collaborations while respecting the privacy of their members.

*Speaker
†Corresponding author: adrian.wuethrich@tu-berlin.de

Keywords: history and philosophy of science, ethics, digital humanities, scientific collaborations, communication

Beyond "Obscurantist Word Play": Deconstruction as an analytical approach in Hans-Jörg Rheinberger's history of life sciences research

Zhuohan Yang * [1]

[1] Goldsmiths University of London – United Kingdom

This presentation explores the linguistic and semantic characteristics of Hans-Jörg Rheinberger's historical epistemological research, with a focus on his unique contribution to the nature of "epistemic things" and "experimental systems" (Rheinberger, 1997). Rheinberger's examination of conceptual frameworks draws on the parallels between linguistic and experimental systems, leveraging Derridean deconstruction as a crucial analytical approach. He frames science as a form of linguistic system, where linguistic theories can be applied. While several reviews (Fruton, 1998; Rangachari, 1998; Thieffry, 1999)highlight his intertwining historical analysis of case-based experimental investigations with epistemological reflections in the laboratory of Paul Zamecnik, few recognize the importance of Derridean thoughts, which some reviewers (Grene, 2002) find "bewildering and in the direction of obscurantist word play". This presentation seeks to highlight the unique and often neglected contributions of Rheinberger's analytical approach, which applies deconstruction from the linguistic realm to the history of science. It moves beyond the perception of such approach as what Grene dismisses as "obscurantist word play".
While Rheinberger positions science as a form of linguistic system, he does not explicitly justify this parallel comparison. This presentation will employ close reading of his work to demonstrate how Derridean concepts like "deconstruction", "traces", and "différance", traditionally applied to semantic systems, can be productively extended to the history of science. Central to Rheinberger's framework are the concepts of "epistemic things" and "experimental systems". He defines "epistemic things" as material entities or processes that constitute the objects of inquiry, and "experimental systems" as the smallest integral working units of research. Rheinberger (1994) argued that a deconstructive approach is necessary for understanding the non-linear narrative of scientific development characterized by "historiality", a term coined by Derrida. Historiality signifies that scientific progress is not a linear, accumulative process, but a series of events driven by "experimental systems" that have their own internal time. This is closely related to "différance", where Derrida emphasizes that meaning is never fully present but is always "deferred" or "traced" back to other meanings, leaving a "trace" of something else in any given word or concept. This presentation argues that a deeper understanding of Derridean thought is crucial for unpacking the essence of these concepts, as their nature cannot be fully grasped without a thorough engagement with Derrida's work.

Next, this presentation will examine a recently published book *Biodeconstruction: Jacques Derrida and the Life Sciences*(Vitale, 2018), a contextualization of Derrida's 1974 seminar *La vie la*

*Speaker

mort (Life Death). Vitale argues that Derrida's philosophical approach to biological life gains explicit value considering contemporary life sciences, which corroborate Derrida's attempts to deconstruct traditional biological frameworks. By shifting the focus to Derrida's own deconstructive interpretation of life sciences, this adds credibility to the application of deconstruction in the history of life sciences.
Finally, this presentation will demonstrate the applicability of deconstruction as a method within the realm of mathematical fluid dynamics. This exploration highlights the potential of deconstruction, as an analytical approach in the history of science, to be applied in fields beyond Rheinberger's application in the life sciences.

Keywords: deconstruction, experimental systems, linguistic systems, historical epistemology

Is Biomimetics a Scientific Discipline?

Dilek Yargan *† [1], Ludger Jansen [1,2], Manfred Drack [1,3]

[1] Institut für Philosophie, Universität Rostock – Germany
[2] PTH Brixen College, Bressanone – Italy
[3] University of Tübingen – Germany

Situated at the crossroads between biology and technology, the areas of biomimetics, biotechnology, and bioengineering – or a more recent approach known as "biological transformation" – develop innovative solutions and products. However, what exactly delineates these research and development fields from each other has not been investigated. In this presentation, we try to state the differences and similarities between biomimetics and its adjacent disciplines and we aim to develop a theoretical foundation for systematising the knowledge generated in biomimetics. The foundation requires that both the epistemology and the ontology of biomimetics are known. As biomimetics seems to have no clearly delineated object of research nor a unified method or objective, our investigation begins with an analysis of biomimetics' epistemic profile, examining its object of research, research approach, methods, objectives, body of knowledge generated, and basic terms. Our general hypothesis is that biomimetics is unique and distinct from other research approaches thanks to these features. We propose a newly developed conceptual framework – comprising function, working principle and construction/design – as beneficial for analysing biomimetics. Using this framework, we propose hypotheses on the epistemic profile and test them by means of case studies. For the case studies, we interviewed researchers involved in biomimetic projects, e.g. in developing blast furnace surface structures based on the water transport system of plants, or in developing sensors based on features of seal whiskers to reduce vortex-induced vibrations. The framework is also analysed ontologically, where a focus is on working principles. Working principles are important in engineering design but, so far, have hardly been investigated by philosophers. The case studies are a basis for ontologically investigating function, working principle and construction/design. Additionally, the framework can be formalised to serve as a tool for a shared way of thinking for all members of biomimetic research teams, thereby ameliorating the tension between scientific goals towards gaining insight into biological systems and the engineers' need to build market-viable applications.

Keywords: Biomimetics, Engineering, Epistemic profile, Ontology, Interviews

*Speaker
†Corresponding author: dilek.yargan@uni-rostock.de

Comparing inaccessible targets through biological modeling: the case of ape iPS cells

Yoshinari Yoshida *† 1

1 University of Exeter – United Kingdom

Philosophical literature on scientific modeling has addressed the problem of how systems or phenomena that are not easily accessible can be studied by employing various modeling strategies. This presentation examines a special version of this problem: when scientists' goal is to *compare* multiple inaccessible target systems, what modeling strategies can be used? To address this question, I focus on a specific example: comparative primatology based on ape stem cell lines. Non-human apes (e.g., chimpanzees, bonobos, and gorillas) are the closest relatives to humans, and hence can be a critical source of information about how human-specific traits (such as the big brain, hairlessness, and bipedalism) have evolved. However, invasive experimental research into them cannot be conducted because of ethical and practical limitations. There have recently been growing attempts to circumvent this problem by employing advanced stem cell technology. It is now possible to create induced pluripotent stem cells (iPSCs), which are expected to be functionally equivalent to embryonic stem cells, from adult tissue samples that are collected from apes in a non-invasive manner. iPSCs are then induced to differentiate into various cell types, which could recapitulate development and physiology of various tissues and organs. My presentation analyzes how ape iPSCs serve as models in this type of research. Unlike model organisms and experimental organisms in biomedical contexts, ape iPSCs are not studied as surrogate models of humans. Instead, each iPSC line is expected to model embryonic development of the species from which it is derived. More importantly, the aim of creating and studying ape iPSCs in this context is not to understand a specific target system, but to *compare* development across the human and other apes to elucidate human evolution. I argue that this comparative goal shapes distinct modeling practices with specific methodological challenges. The ideal of experimental control suggests that scientists should use the same culture protocols to generate and manipulate iPSC-derived models to be compared; even if iPSC-derived models are distorted from the targets, as long as they are distorted in the same way, differences between the models can be attributed to differences between the targets. At the same time, iPSC-derived models are highly variable across species and across individuals, which means that a culture protocol that works for one model does not necessarily work for another model. These considerations lead to a peculiar methodological challenge: there is a trade-off relationship between a precise inter-model comparison and the quality of individual models. Furthermore, this methodological challenge suggests a more basic epistemic problem: if the target systems might differ in developmental time and internal bodily environments, what does it mean to treat their models in the same way for a precise comparison? I characterize this problem by appealing to the philosophical literature on experimental control and discuss potential solutions.

*Speaker

†Corresponding author: y.yoshida@exeter.ac.uk

Keywords: Stem cells, in vitro models, evolutionary developmental biology, comparative biology, experimental control

Philosophy of Science Meets HRI: Towards a Critical Epistemology of Robotic Emotions

Ivano Zanzarella *† [1]

[1] Università degli studi di Bari Aldo Moro = University of Bari Aldo Moro – Italy

Today, Human-Robot Interaction (HRI) is a well-established research field, distinguished by its strong interdisciplinary focus that brings together engineering, artificial intelligence, psychology, and neuroscience. However, as a young scientific discipline, it has just recently begun to attract the attention of philosophers of science, to the effect that many aspects of its underlying epistemology have never been questioned or (normatively) articulated until now, in comparison to other ("elder") scientific disciplines.
The principal aim of my proposal is to contribute to establishing a more precise epistemology for HRI, here in particular with respect to the issue of robotic emotions and emotional attributions to robots.

HRI, in fact, is currently studying how the attribution of mental states, and specifically emotions, to robots works. In most cases, however, the (empirical) research in the field failed to recognize the complexity of emotional phenomena as displayed by philosophers and cognitive scientists in at least the last two centuries. Emotions have been often indistinctly subsumed and investigated under the general category of "mental states" together with beliefs, knowledge, desires, intentions, etc. Moreover –from within an implicit and mostly uncritical acceptance of a reductivist (and non-intentionalist) theory of emotion– the emergence and the attribution of emotional states have been mostly correlated only with physical and bodily factors, without sufficient attention to the other phenomenological, cognitive and expressive-behavioral elements however constitutive of those particular states. Finally, the conception of emotion considered by robotics and HRI does not seem to refer to evaluative and propositional components involved in emotional expression and attribution, to the effect that, for example, attributing an emotion to a robot also means attributing in part some sort of propositional attitude (e.g. a belief) to it.

Thus, firstly, it emerges the need for a systematic mapping, in current scientific literature, of all the different conceptions of emotion (often implicitly) taken as theoretical basis both in robotics for the development of emotional robots and in HRI for studying human attributions of emotional states to them. Secondly, from considering the shortcomings of applying naïve emotion theories in robotics and HRI, the demand has to be addressed of a normative delineation of a comprehensive and critical epistemology of emotions which is suitable both for developing emotional robots and to reflect about how humans attribute emotions (and which kind of emotions) to robots.
Relying on the traditional methodology of philosophy of science, my proposal aims to respond to these urges, which is relevant also considering the fact that robots and machines are currently

*Speaker
†Corresponding author: ivano.zanzarella@uniba.it

gaining an ever growing importance in many areas of human life and society. In this sense, examining the epistemology underlying robotics and HRI, avoiding possible bias and trying to give solutions to compelling epistemological issues is significant also for favoring the future design of technologies which are epistemically more transparent and morally more trustworthy (and empathetic).

Keywords: Human, Robot Interaction, Robotics, Emotions, Epistemology of AI

The Benchmarking Epistemology: What scientists can learn from competitive comparisons of prediction models

Sebastian Zezulka *† [1], Timo Freiesleben *

1

[1] Universität Tübingen – Germany

It is standard methodology in machine learning (ML) to evaluate the quality of a model based on its predictive performance on a test dataset and rank it against other competitive models – a practice known as benchmarking. Benchmarking is integral to ML research papers and has been coined the iron rule of machine learning. As ML becomes more widespread in the natural and social sciences, benchmarking is also increasingly adopted as an evaluation standard of scientific significance.
While benchmarking is ubiquitous both in ML and in scientific applications it has so far drawn little attention from the philosophy of science community. This paper argues that benchmarking is a scientific methodology in its own right. It shows that benchmarks can support various scientific inferences, for example:
• Choosing a model to deploy on a specific task.
• Assessing the predictability of a target Y from features X.
• Comparing empirical adequacy of different scientific theories or models.
• Inferring the most suitable inductive learning strategies for a specific class of tasks.

However, for each of these inferences, additional assumptions must be made that link the studied phenomenon and the respective scientific problem to the operationalization of the prediction task, the evaluation metric, and the representativeness of the benchmark data.

We illustrate our argument and the specific assumptions required with three case studies from different domains: the Imagenet benchmark for image recognition in computer vision, the fragile families benchmark for predicting life outcomes in the social sciences, and the WeatherBench benchmark for global weather forecasting in ecology.

Next to its epistemic role, benchmarks also have an important social role in organizing scientific communities. The benchmarking methodology offers:
• a well-defined prediction task including the data to tackle it,
• a shared concept and measure of progress,
• and a shared language to communicate results.

*Speaker
†Corresponding author: sebastian.zezulka@uni-tuebingen.de

The benchmarking epistemology faces several limitations, particularly in combination with its social role. The iterative use of benchmark data to rank models breaks standard guarantees from statistical learning theory. Like p-values, benchmarks are often used as a standard of scientific significance in peer review processes, incentivizing gaming the benchmarking metric or offering incremental performance gains over the contribution of genuine scientific insights. Moreover, benchmarks can easily be misinterpreted in their scope by reducing model evaluation to a single number: The insights benchmark challenges can offer are tied to the definition of the task, given in the evaluation metric and the data used. Scientific questions that do not concern predictive accuracy but causal explanations or problems that are more difficult to operationalize and measure cannot be straightforwardly addressed using benchmarks.

Given these limitations, we discuss how benchmarks can license inferences for deployment, assess predictability, compare the empirical adequacy of models, and infer suitable learning strategies.

Keywords: Model Evaluation, Benchmarking, Machine Learning, Scientific Modeling, Prediction, Philosophy of Science

Author Index