MCMP-Wuppertal-Hannover Workshop

Location: Edmund-Rumpler-Strasse 9 A 011
The closest U-bahn is Freimann (U6)

Thursday 6th July 2023

9.30 - 10.20	John Dougherty (MCMP)	Chair: Alice Murphy
10.30 - 11.20	Luis Lopez (Hannover)	
11.30 - 12.20	Daniel Minken (Wuppertal)	
	Lunch	
13.00 - 13.50	Leon Assaad (MCMP)	Chair: Martin King
14.00 - 14.50	Radin Dardashti (Wuppertal)	
15.00 - 15.50	Donal Khosrowi (Hannover)	

16.00: Drinks at Atzinger

18.00 - 20.00: Public evening lecture by Sabine Hossenfelder

20.15: Workshop dinner at Kun-tuk

Friday 7th July 2023

9.30 - 10.20	Anna Leuschner (Wuppertal)	Chair: Mathias Frisch
10.30 - 11.20	Gábor Hofer-Szabó (MCMP)	
11.30 - 12.20	Ina Gawel (Hannover)	
	Lunch	
13.00 - 13.50	Rafael Fuchs (MCMP)	Chair: Alice Murphy
14.00 - 14.50	Leon-Philip Schäfer (Wuppertal)	
15.00 - 15.50	Birgit Benzing (Hannover)	

Titles and Abstracts

John Dougherty: **Constructive empiricism and the scientia mensura**

Wilfrid Sellars's scientific realism is memorably captured by his scientia mensura: "in the dimension of describing and explaining the world, science is the measure of all things". I argue that in The Scientific Image, van Fraassen exploits a tension between description and explanation in this apothegm. Read this way, van Fraassen's arguments against Sellarsian realism are more difficult to dismiss than is usually supposed, and stock objections to constructive empiricism lose some of their bite. This understanding of constructive empiricism also bears on current debates over the compatibility of Sellars's scientific realism with the rest of his philosophical system and over the status of semantics in realism.

Luis Lopez: **Genuine Understanding or Mere Rationalizations? Approximations and Idealizations in Science and Explainable AI**

Rudin (2019) has prominently argued that local post hoc XAI models are inherently misleading, as they offer mere rationalizations, rather than explanations, of decisions made by black box machine learning models. In response, some philosophers of science have been too quick to construe that these arguments stem from a normative premise according to which perfect faithfulness between (local) post hoc XAI models and their targets is necessary for genuine understanding. Moreover, they have been even quicker in drawing insights from the literature on idealized scientific models to challenge such a premise. I show how these responses not only mischaracterize what is at the core of Rudin’s arguments but also fail to distinguish idealization from approximation. I argue that local post hoc XAI models are inherently misleading due to approximation failure, rather than imperfect faithfulness (Fleisher, 2022) or idealization failure (Sullivan, unpublished).

Daniel Minken: **Experts: An application-oriented approach**

The questions of what a scientific expert is and whether we should place unrestricted trust in experts has been debated in many disciplines for more than 50 years. The concept of expertise has also become a research topic in Applied, Social, and Political Epistemology. Here, different authors have offered various definitions of this concept and pointed out the enormous importance of applying theoretical groundwork to societal controversies (e.g., about anthropogenic climate change, the use of alternative medicine, etc.). However, many of the application-oriented approaches in these areas remain, from my perspective, too general. In my talk, I would like to make them more concrete by connecting the aforementioned theoretical groundwork with a new application example from the field of artificial intelligence.

In the first part, the phenomenon of expert disagreement is examined by, among other things, attempting to explain the concept of expert adequately. To this end, I will survey the newer expertise debate in epistemology and criticize influential approaches. In a second step, I will describe the theoretical foundations of a Machine Learning System—the so-called "Automatic Deception Detection System" (ADDS)—that was developed to detect illegal entry into the

European Union. These foundations will serve as the central application example. The last part will argue that the said foundations of the ADDS are based on ignoring a case of expert disagreement. My main claim will be that the development of ADDS was not justified in light of epistemological considerations, which will demonstrate the potential of Applied Epistemology and Philosophy of Science to contribute to important societal issues of our time.

Leon Assaad: **Polarization and radicalization among Bayesian agents: when sharing and communicating only worsens disagreement**
Public discussions of relevant social, political or scientific issues sometimes lead to persistent polarization and disagreement. Some prominent accounts treat belief polarization as the result of epistemic irrationality. Recently, however, formal epistemologists have shown that polarization can arise even among rational agents under subpar conditions. Here, I draw on this formal literature to construct a simulation model of Bayesian agents aiming to determine whether a proposition is true. They share full access to the same set of evidence and communicate about how that evidence should be evaluated. Perhaps surprisingly, polarization can occur even under such favorable conditions. The key mechanism generating polarization involves agents evaluating the evidence differently based on different background beliefs. When this is the case, rather than being a remedy, direct communication can exacerbate polarization by radicalizing the agents' background beliefs. This finding is reminiscent of analyses of political polarization and "deep disagreements" in contemporary societies, and suggests that such problematic phenomena may be rational and worsened by simple social dynamics.

Leon-Philip Schäfer: **Ad Hominem Arguments in Scientific Discourses**
Ad hominem arguments have a peculiar status within scientific reasoning that seems to be surprisingly unfathomed in modern philosophy of science. Therefore, I would like to draw our attention to this interesting topic, by examining the significance of ad hominem arguments in the context of scientific discourses. In particular, I would like to explore whether ad hominem arguments are genuinely fallacious inferences that should be kept out of scientific discussions altogether; or whether we should embrace a more nuanced evaluation of ad hominem arguments that allows for the view that such arguments can be legitimate sometimes. The claim I would like to defend says that certain ad hominem arguments, like the subtype poisoning the well, are problematic in the context of scientific discourses because they can be used as immunisation strategies, i.e. as tools which allow us to protect our theories against any possible kind of criticism such that rational discussion is made impossible.

Donal Khosrowi: **Can AI produce synthetic evidence?**
The recent proliferation of generative artificial intelligence systems (GAI) raises a host of novel philosophical questions about AI in science. We focus on the role of GAI in the historical sciences, including history, archaeology and anthropology. Here, researchers are already using AI for various purposes, e.g. to reconstruct partially destroyed manuscripts, and a new wave of GAI systems like StableDiffusion hints at the possibility of more ambitious restorative inferences. For instance, suppose researchers trained a GAI model on extensive labeled image, text and scan data on artefacts recovered from a certain region. Consider now a case where researchers prompt such a system to provide a rendition of how a newly discovered, but partially destroyed

artefact would have looked like if it had remained intact. Can we sometimes consider such outputs to be epistemically on par with expert judgment or finding concrete, material evidence speaking to the same query? A first blush response is to say no: GAI outputs may be understood as hypotheses or speculations, and there might be good reasons to pursue these hypotheses, but they are not evidence in and of themselves. Contra this view, we argue that under suitable conditions, e.g. concerning the nature, amount and variety of training data, the fidelity of theory that informed the labeling of these data, and constraints on learning processes, GAI outputs can indeed carry enough derivative and intrinsic justification to count as synthetic evidence, which can sometimes be epistemically on par with traditional forms of evidence, e.g. material evidence or expert judgment. We show how our thesis connects with prior arguments in the epistemology of computer simulation and modeling, and explore the wider ramifications it has for the epistemology of AI-based science.

Anna Leuschner: **The Strategic Intimidation of Scientists**

Scientists working in fields that threaten powerful industrial or political interests have become increasingly strategically attacked. These attacks fall largely into two categories: professional attacks, where scientists' work is being marginalized and their competence discredited, and personal attacks, where they are overtly bullied, harassed, and ridiculed. In this talk, I will explore the strategies behind these attacks: the so-called Tobacco Strategy, which encompasses the so-called Serengeti Strategy and the Manufacture of Dissent, as well as a rather new strategy that we wish to call the Cancel Culture Strategy.

While personal attacks are, in most cases, obviously inadmissible, the case is more difficult when it comes to professional attacks as the parties behind these attacks invoke the freedom of speech: in particular the strategic Manufacture of Dissent and the Cancel Culture Strategy are difficult to counter; this requires thoughtful analyses on a case by case basis, where special interests, biases, and power relations have to be taken into account.

Gábor Hofer-Szábo: **On the three types of Bell's inequalities**

Bell's inequalities can be understood in three different ways depending on whether the numbers featuring in the inequalities are interpreted as classical probabilities, classical conditional probabilities, or quantum probabilities. In the talk I will argue that the violation of Bell's inequalities has different meanings in the three cases. In the first case it rules out the interpretation of certain numbers as probabilities of classical events. In the second case it rules out a common causal explanation of conditional correlations of certain events (measurement outcomes) conditioned on other events (measurement settings). Finally, in the third case the violation of Bell's inequalities neither rules out the interpretation of these numbers as probabilities of events nor a common causal explanation of the correlations between these events---provided both the events and the common causes are interpreted non-classically.

Ina Gawel: **Considering early literature on peer review: why the current discussion on peer review would benefit from an HPS approach.**

The subject of peer review tends to be controversial, and has recently been the subject of some sensational claims: Heesen and Bright (2021), for example, called for the abolition of peer review. Drastically written (or emotionally justified) texts about peer review are not a new phenomenon, even if this does not make the discussion any more fruitful. In fact, I would argue that the discussion about peer review - in its current form - is not fruitful at all. In this talk, I would like to try to bring some clarity to the discussion by focusing on three questions: first, why we are talking about peer review at all and how this discussion came about; second, what we are actually talking about when we talk about the discussion of peer review; and third, why no solution is possible in this supposedly solution-focused debate. I then go on to explain why I advocate a HPS approach to the debate, which should be prioritised over any further discussion of peer review.

Rafael Fuchs: **An agent-based model for assessing probabilistic measures of Coherence**
Coherence is an important epistemic value that figures in theory development, argumentation, and many other epistemic activities. However, it is still unclear how coherence should be measured. The literature on probabilistic measures of coherence has developed a large variety of candidates, and their competition continues. To assess probabilistic measures of coherence in terms of their truth conduciveness, I propose an agent-based model, in which a detective must find the true suspect by assessing testimonial statements using different measures of coherence. The model allows for a direct comparison of various candidate measures in terms of singling out the correct suspect, and on top of that, the measures can be compared against a Bayesian benchmark. Primary results suggest that correlation-based measures emerge as strong contenders.

Radin Dardashti: **On whats and thats in what-that and that-what discoveries?**
Schindler (2015) has defended and clarified Kuhn's account of scientific discovery. According to him, scientific discoveries come in two forms: "that-what" and "what-that" discoveries. The former involves discovering the existence of something already known, while the latter involves first discovering that something exists and then determining what it is. However, the distinction between these types of discoveries is conceptually problematic. In this talk, I suggest that the what-part of a discovery is better understood as a gradual process rather than a categorical distinction. Moreover, any that-discovery already presupposes some degree of what-discovery. I will argue on the basis of a systematic analysis as well as on the basis of examples from the history of physics.

Birgit Benzing: **Epistemic Repercussions on Socially Mandated Research**

Abstract: Animal welfare science has been called a mandated science, i.e. a science that originated not from curiosity but from ethical concerns of society. I call this “the narrative of mandated science.” The details of the narrative have shifted over time, however. In this paper, I connect these shifts and transitions to different roles the narrative serves and different audiences it addresses. Until now, the narrative has been debated in the fields of animal rights / liberation and Human Animal Studies (HAS), situating it in a moral and political context. However, these debates neglect the epistemic repercussions of the narrative, which becomes more apparent in the most recent version of the narrative. My paper will lay out these epistemological repercussions and argue that AWS studies have to confront the fact that their social concern are not external to their epistemic endeavours.