Book of Abstracts

International Conference on Engaging Ethics and Epistemology in Science (EES 2022)

29-30 September 2022
Leibnizhaus Hannover

Leibniz Universität Hannover

DFG Deutsche Forschungsgemeinschaft German Research Foundation

UNIVERSITÄT BIELEFELD

The conference is organised by DFG-Graduiertenkolleg 2073 “Integrating Ethics and Epistemology of Scientific Research” and funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project 254954344/GRK2073/2.

Abstracts

The International Conference on Engaging Ethics and Epistemology in Science

29-30 September 2022 · Leibnizhaus Hannover, Germany

The International Conference on Engaging Ethics and Epistemology in Science (EES 2022) focuses on bridging the gap between epistemological and ethical or political approaches as well as theoretical and material approaches to philosophy of science. Many important questions concerning the role of science in society and its ability to contribute to policy debates require integrating political, ethical, material, and epistemological perspectives on science.

The conference includes contributions from fields such as:

- *Values in Science*: Under what conditions do non-epistemic values legitimately influence scientific and technological research? How can we adjudicate the moral or political legitimacy of non-epistemic value influences in particular cases of research?
- *Social Epistemology of Science*: What is the best arrangement of scientific communities for knowledge production? Are different arrangements best for different epistemic goals? How does social diversity among scientists influence epistemic practices?
- *Science in a Democratic Society*: How do the principles of democratic and scientific institutions complement or conflict with each other? Ought scientific research be participatory (i.e., 'citizen science' or 'community science')?
- *Trust in Science and Science Communication*: What form of trust between scientific communities and public groups are required for science in a democratic society?
- *Socially Responsible Research and Innovation*: What kind of ethical problems are posed by scientific and industry research? How should dual-use research be evaluated and regulated?
- *Uncertainty and Explainability*: How do values play a role in mitigating or reducing epistemic uncertainty? Does increasing the explainability of scientific models reduce uncertainty?
- *Objectivity in Science*: Ought science be objective and in what sense? By what procedures can we improve or evaluate the objectivity of science?
- *Philosophy of Experimentation*: What epistemological strategies are useful in concept formation and the development of new methodologies? How does the material culture of science impact the construction of models or assessment of theories?

The conference consists of parallel sessions integrating talks by invited speakers and those who have submitted papers. We are very pleased that the following invited speakers have accepted our invitation: Kristen Intemann, Tarja Knuuttila, Sabina Leonelli, Anna Leuschner, Liam Kofi Bright, and Kevin Elliott.

Location

This is the address of our conference location:

Leibnizhaus Hannover, Holzmarkt 4–6, 30159 Hannover, Germany

The original Leibnizhaus was a Renaissance townhouse in Hannover built in 1499 and named after the philosopher Gottfried Wilhelm Leibniz, who is also the namesake of the University of Hannover. Leibniz lived in the house from 1698 until his death in 1716.

During World War II, the building was destroyed in an air raid on Hannover in 1943. Between 1981 and 1983, a reconstructed new building with the faithfully reproduced facade was erected at Holzmarkt in Hannover.

General Schedule

Please see https://grk2073.org/eesconference2022/schedule/ for the detailed schedule.

Day 1 – Thursday, 29 September 2022

08:15-08:45 **Registration**

08:45-08:55 **Welcome Address (Organising Commitee)** *Leibnizsaal, Ground Floor*

08:55-09:15 **Welcome Address (Prof. Dr. Torsten Wilholt)** *Leibnizsaal, Ground Floor*

09:20-10:00 **Invited Talk by Kristen Intemann** *Leibnizsaal, Ground Floor*

10:05-10:45 **Parallel Sessions** *Leibnizsaal (Ground Floor), Karmarschraum (1st Floor), and Kerstingzimmer (2nd Floor)*

10:45-11:15 **Coffee Break** *Hannah-Arendt-Saal, Basement Floor*

11:15-11:55 **Parallel Sessions** *Leibnizsaal (Ground Floor), Karmarschraum (1st Floor), and Kerstingzimmer (2nd Floor)*

12:00-12:40 **Invited Talk by Anna Leuschner** *Leibnizsaal*

12:40-14:05 **Lunch** *Hannah-Arendt-Saal, Basement Floor*

14:05-14:45 **Parallel Sessions** *Leibnizsaal (Ground Floor), Karmarschraum (1st Floor), and Kerstingzimmer (2nd Floor)*

14:50-15:30 **Invited Talk by Liam Kofi Bright** *Leibnizsaal*

15:30-16:00 **Coffee Break** *Hannah-Arendt-Saal*

16:00-16:40 **Parallel Sessions** *Leibnizsaal (Ground Floor), Karmarschraum (1st Floor), and Kerstingzimmer (2nd Floor)*

16:40-16:50 **Short Break**

16:50-17:30 **Parallel Sessions** *Leibnizsaal (Ground Floor), Karmarschraum (1st Floor), and Kerstingzimmer (2nd Floor)*

17:35-18:15 **Invited Talk by Kevin Elliott** *Leibnizsaal*

19:30-21:30 **Conference Dinner** (at own cost), *Brauhaus Ernst August, Schmiedestraße 13, 30159 Hannover*

Day 2 – Friday, 30 September 2022

08:45-09:00 **Welcome & Announcements** *Leibnizsaal*

09:00-09:40 **Invited Talk by Tarja Knuuttila** *Leibnizsaal*

09:45-10:25 **Parallel Sessions** *Leibnizsaal (Ground Floor), Karmarschraum (1st Floor), and Kerstingzimmer (2nd Floor)*

10:25-10:55 **Coffee Break** *Hannah-Arendt-Saal*

10:55-11:35 **Parallel Sessions** *Leibnizsaal (Ground Floor), Karmarschraum (1st Floor), and Kerstingzimmer (2nd Floor)*

11:40-12:20 **Invited Talk by Sabina Leonelli** *Leibnizsaal*

12:20-12:40 **Closing** *Leibnizsaal*

12:40-14:00 **Lunch** *Hannah-Arendt-Saal*

14:00-16:00 **Afternoon Hike**

Thursday, 29 September 2022

09:20-10:00

Kristen Intemann:
Science and Values: Understanding Distrust of Predictive Models (Invited Talk)

Leibnizsaal

Maya Goldenberg (2022, forthcoming) argues that warranted trust requires that scientists and scientific institutions demonstrate i) epistemic competency, ii) moral reliability, and iii) a commitment to the public interest. It is often assumed that public skepticism of predictive models (e.g. climate change impacts or COVID-19 models) is based on false beliefs about how models work or the reliability of predictions. Thus, modelers focus on how to improve the actual or perceived epistemic competency of models. I argue that there are a host of other social and ethical reasons why various communities distrust predictive models. These include distrust of how models are being used, concerns about what is (and is not) represented in models, lack of acknowledgement of the ways that values play a role in both modeling and policy decisions, and an awareness of the ways in which certain models can be (unintentionally) biased. Thus, addressing the lack of trust in modelling will require not just increasing the reliability or epistemic competency of models, but also developing models and communicating about them in ways that can promote moral reliability and wellbeing. How best to do this is considered, particularly given concerns (e.g. John 2018) that having scientists be open, transparent, honest, and sincere may actually exacerbate unwarranted distrust among those whose trust in experts is already "fragile".

10:05-10:45

Deniz Sarikaya:
A Norm for Science Advice: Making Beliefs Accurate

Leibnizsaal

How should scientists communicate their findings when they advise politicians? One view holds that scientists should say what they have a high credence in. For example, they should not assert 'X is toxic' if they only have a credence of 0.7 that X is toxic. Rather, they should make their uncertainty explicit to say something weaker that they have a high credence in, such as 'it is likely that X is toxic'. Another view holds that scientists should say what they expect to have the best policy consequences. For example, if scientists know that politicians will enact climate policies only if scientists do not make their uncertainty explicit, and the scientists think that climate policies are desirable, then they should not make their uncertainty explicit. We explore a third view, according to which scientists should say what they expect to make the politicians' credences most accurate. That is, if a scientist has a credence of 0.7 in X being toxic, then she should say whatever brings the politician's credence close to 0.7. If this requires not making uncertainty explicit or saying things she takes to bring about suboptimal policy consequences, so be it. For ease of reference, let us state the three views as three norms for scientists advising politicians.

(Honesty) Advising scientists ought to say what they have a high credence in.

(Policy) Advising scientists ought to say what maximises the expected value of the policy consequences of what they say.

(Addressee) Advising scientists ought to say what maximises the expected accuracy of their addressee's credences in the target propositions.

This talk explores the advantages, problems, and implications of (Addressee). First, we outline potential advantages of (Addressee) over its two alternatives. In particular, (Addressee) does not permit scientists to skew their advice based on their moral assessment of policies. This is an advantage over (Policy) because such influence would undermine procedural values of democratic decision-making. It also constitutes a defense of the value-free ideal, which is sometimes attacked on the basis of a norm such as (Policy). Also, (Addressee) does not require scientists to say what they have a high credence in even if that is counterproductive to induce more accurate credences in the addressee. This is an advantage over (Honesty) because, in such cases, it seems permissible to say what one has a low credence in, if that makes the politician's credences more accurate. We then turn to a central problem of (Addressee), that it seems to require vicious communication strategies if those happen to maximise the expected accuracy of the politician's credences. Finally, we assess prominent examples of science advice in the light of (Addressee).

Just as many other activities, science advice might be governed by different kinds of norms which employ different senses of 'ought'. Maybe there is a moral norm, which tells us what scientists morally ought to say, and an epistemic norm, which tells us what scientists epistemically ought to say. Then, the above norms might not be mutually exclusive: they might simply employ different senses of 'ought'. But we read all these norms in a specific sense of 'ought': the moral sense. What we are after is a moral norm of science advice, not an epistemic norm.

Giorgia Pozzi:
Machine Learning-induced Epistemic Injustice in Healthcare: Epistemological and Ethical Perspectives

Karmarschraum

Artificial intelligence-based (AI) systems are playing an increasingly relevant role in the field of medicine and healthcare, bringing about novel ethical and epistemological issues that need to be timely addressed. Since AI systems are epistemically authoritative, hardly contestable, and increasingly involved in decision-making procedures that strongly impact patients' lives, it is fundamental to make sure that they do not undermine epistemic subjects in their capacities as knowers, i.e. as recipients and conveyers of knowledge. Through the analysis of the shortcomings of an ML system currently deployed in the USA to predict patients' likelihood of opioid addiction and misuse (PDMP algorithmic platforms) (Szalavitz 2021; Oliva 2021), in this presentation, I point at forms of epistemic injustice (Fricker 2007) brought about by the use of opaque AI systems in the healthcare domain. More concretely, drawing on the case of PDMP systems, I aim to spell out how hermeneutical injustice can be AI-induced. I show that this is the case through three argumentative steps. First, I show that the AI system involved in decision-making processes in the case under scrutiny holds an unwarranted hermeneutical privilege. This privilege amounts to the fact that the system plays a decisive role in shaping shared hermeneutical resources (i.e. meanings and concepts). It is unwarranted due to the fact that it eludes human intervention and control.

Relatedly, in a second step, I show that the way in which hermeneutical resources are established hinders understanding and communication among relevant stakeholders (i.e., patients and physicians), rendering significant aspects of their social experience not intelligible to them. Third, the interplay of the factors previously mentioned leads to considerable disadvantages, particularly at the expense of socially marginalised groups. These disadvantages manifest in epistemic, moral, and practical terms. I further argue that AI-induced hermeneutical injustice is particularly harmful due to what I define as an automated hermeneutical appropriation that occurs from the side of the AI system. This happens because the system does not provide physicians with meaningful explanations, but they are, in turn, directly affected by the results of the decision-making process itself (e.g. to the extent that they are expected to take up epistemic and moral responsibility for the decision taken). This creates an epistemic dependence that undermines their epistemic authority and may lead to a proper injustice. Crucially, this strongly constrains also epistemically well-positioned agents in their possibility to mitigate possible cases of hermeneutical injustice experienced by vulnerable agents through virtuous behaviour (Fricker 2007: 174). With my presentation, I hope to shed light on a present concern that needs to receive due attention. More generally, my presentation should expand the analysis of ethical issues raised by AI systems that are to be considered epistemic in nature, thus bridging the gap between these two dimensions in the ongoing AI debate.

Joeri Witteveen:
Ontic Risks, Uncertainty, and Classification for Biodiversity Conservation

Kerstingzimmer

This talk has two aims. The first aim is to *reconfigure* ongoing debates about the nature of inductive risk versus other kinds of epistemic risks. After Heather Douglas successfully put the argument from inductive risk back on the philosophical agenda (Douglas 2000, 2009), Justin Biddle and Quill Kukla argued that we ought to distinguish different kinds of risk in science – inductive risk being one of them – that fall under the umbrella category of epistemic risk (Biddle 2016, 2020; Biddle & Kukla 2017, Kukla 2019). More recently, Stephanie Harvard and Eric Winsberg (2022) have argued that the literature on epistemic risk lacks a systematic framework for distinguishing between risks within this broader category. They argue that the non-inductive risks that Biddle, Kukla, and others have identified can be unified under an account of representational risk. On Harvard and Winsberg's account, representational risk is the only kind of risk that "truly rivals" inductive risk in significance. I argue that Harvard and Winsberg's account is problematic, since it hinges on an interpretation of risk that fails to do justice to the need to "balance risks" that is recognised in the standard literature on inductive risk. However, I argue that some accounts of (non-inductive) epistemic risk – in particular those that focus on classification in the biomedical sciences – do adequately recognise this property of "risk taking" in scientific practice. I therefore suggest that a systematic taxonomy of risks in science should distinguish between inductive and classificatory risks.

The second aim will be to *extend* this taxonomy of risk in science by arguing that it fails to recognise the presence of ontic risk in certain areas of policy-relevant science. The case for the notion of ontic risk is developed through an example from biological taxonomy at the boundary with species conservation policy. The motivating thought behind the notion of ontic risk is that classificatory choices at the interface between biological taxonomy and species conservation policy can be performative in an ontological sense. Classificatory choices can differentially shape the "ontological futures" of taxonomic groups, through the anticipated effects of conservation actions. The point here is not the mundane one that conservation actions can save species, but rather that they can reinforce species circumscriptions through their choice of species concept. This leads to an analysis of risk that is distinctly ontological in orientation and cannot be reduced to either inductive or epistemic risk.

11:15-11:55

Vincent Cuypers:
A Case for Voting in Science

Leibnizsaal

In science, much effort is often required to solve issues of conceptual and classificatory organisation. For science to be efficient as a collective enterprise, agreement needs to be found on which concepts and classifications to use in representing reality. However, such concepts and classifications are often subject to substantial debate. For example, much of biology is organised around the notion of species, but disagreements abound on what species actually are, and on which groups of organisms must be recognised as species. Likewise, in medicine there are fierce debates on the recognition and classification of diseases and disorders, and in geology there is debate on the recognition of the Anthropocene as a separate geological epoch. These issues have in common that they require a solution with a certain urgency, because the concepts and classifications concerned constitute an important input in scientific research; that it matters which solution is eventually chosen, as concepts and classifications impact the research in which they are used; and that they often arouse much emotion, among scientists and the broader public alike. One method that seems to be more and more in vogue to efficiently resolve such disagreements is to put them to a vote. For example, there have been votes on the derecognition of homosexuality as a mental disorder, on the definition of planets and correspondingly the status of Pluto as a dwarf planet, and on the issue of the Anthropocene. Similarly, in biology, the potential of voting on species classifications is explored by some working groups as a means to reduce taxonomic disagreement. However, these votes are often controversial, and their outcome is rarely accepted without resistance, showing that reflection is required on how such voting should be organised, and how its legitimacy can be assured.

Here, we investigate whether and how voting can indeed be a legitimate way to solve conceptual and classificatory controversies. For this, we explore current examples of such votes, and draw lessons to develop a framework according to which voting on conceptual and classificatory issues can be theoretically sound and practically workable. We focus in particular on the role of non-epistemic values in guiding conceptual and classificatory decisions, and the question whether such value-judgements can be made on a collective level through voting. Also, we focus on some problems, including the question of who may vote, and the question of how collective conceptual and classificatory agreement can be reconciled with academic freedom.

Jon Leefmann:
Building Public Trust in Science: Why Science-Advertising is a Bad Idea

Karmarschraum

An important distinction in the philosophy of trust concerns the question of whether or not the attitude of trust necessarily involves the belief that the trusted are trustworthy. In accounts that respond to this question affirmatively, warranted trust depends on epistemic reasons. On most of these accounts, in order to adopt the attitude of trust one needs evidence in support of the belief that the trusted are trustworthy. Accounts of public trust in science typically presume this picture of trust and have good reasons to do so. Yet, I will argue, it has some remarkable implications for the possibility to build trust in science. If trust depends on evidence in support of the belief that the trusted is trustworthy, trust cannot be gained by presenting oneself as trustworthy. Self-advertising is a bad idea. The same goes for science advertising: Trust in science is not built by pointing out how trustworthy (i.e. successful, competent, and socially responsible) it is. This will not be seen as evidence of trustworthiness but as “doctored evidence” (Moran) presented with the intention to make believe. The fact that one presents something as evidence to someone taints the evidence. Self-ascriptions of trustworthiness will provoke skepticism rather than trust.

As a consequence, I will argue that attempts to build trust in science should rather focus on the barriers that hinder the public to perceive science as trustworthy. I will discuss three points of departure: (1) an actual lack of trustworthiness, (2) a distorted public discourse, and (3) emotional attitudes that influence the capacities of members of the public to become aware of relevant evidence for trustworthiness. With respect to the first point, I will turn to the recent debate on scientific objectivity to identify practices in science that undermine its trustworthiness. I will identify competence, sincerity and social responsibility as important dimensions of scientific trustworthiness and discuss how science could enhance its social responsibility through becoming more diverse and more inclusive. With respect to the second point, I will urge to reconsider the norms of properly communicating scientific knowledge to the public. I will suggest a procedure how the public communication of scientific results can be balanced to adjust on the one hand to the anticipated reception of testimony and to avoid the advertising of science on the other. The last point is the most difficult one. Emotional attitudes often have a long history and cannot be influenced directly. I will conclude, however, that it will already be a first step for building trust if their existence is properly acknowledged instead of being discredited as irrational.

Oliver Buchholz and Karoline Reinhardt: How to Use Explainable AI Responsibly

Kerstingzimmer

The increasing relevance of AI systems paired with their repeatedly observed opacity gave rise to the field of explainable artificial intelligence (XAI). XAI tries to overcome said opacity, thereby reducing negative societal impacts of and fostering trust in AI systems. However, XAI itself raises several important epistemological and ethical questions. On the one hand, its epistemic goal is unclear. Proposals range from *explainability* to *interpretability* (Erasmus et al. 2020) or *understandability* (Páez 2019). On the other hand, there is a wide range of techniques to achieve these goals, ranging from local approximations of complex models (Ribeiro et al. 2016) to visual (Xu et al. 2015) or textual explanations (Hendricks et al. 2016). Furthermore, even the relation between goals and techniques is debated extensively. Finally, some argue the epistemic goal only serves as a means to ultimately achieve an ethical end: Employing XAI helps to avoid discrimination (Krishnan 2020) or allows individuals to contest adverse decisions (Wachter et al. 2018).

What is commonly sidestepped, however, is the distinction between the *ability* to explain and the *obligation* to explain. Usually, it is assumed rather vaguely that an explanation should be provided for consequential decisions of AI systems (Lipton 2018), when their safe use is crucial (Doshi-Velez & Kim 2017), or in contexts in which the stakes are generally high, for instance in the medical field (Watson et al. 2019). However, this leaves important questions unanswered: In which *specific* cases and under what *specific* circumstances do we have an obligation to give an explanation? And following from that: In which cases and under what circumstances is it necessary that an AI system provides an explanation? This omission is problematic for at least two reasons: On the one hand, it remains unclear which characteristics of a situation establish an obligation to explain; on the other hand, it also leaves open what an explanation should look like that achieves the goals associated with XAI.

In our paper we shed new light on the relation between the ability to explain and the obligation to explain in the context of XAI by combining approaches from moral philosophy, political philosophy, and epistemology. After an introduction (§1), we distinguish different types of explanation and outline why explanations can be ethically relevant (§2). Building on Kantian and Neo-Kantian theories we then provide a framework that systematises in which instances there is a (moral) obligation to explain (§3). The "right to justification" (Forst 2007) is one focus point which we will distinguish from a "right to explanation" as codified in the GDPR. We also look at instances in which individuals have a right to be informed without this amounting to a full-blown right to justification. Finally, we explain why there are also instances in which there is a (moral) obligation to explain though nobody holds a "right" to either explanation or justification. Our systematisation reveals that there are certain cases in which there is no obligation to provide an explanation, and that it is sometimes even (morally) forbidden to provide one (§4). The systematisation developed will subsequently also help to clarify what kind of explanation is needed – if at all. Thus, the paper not only closes an important gap in clarifying when we need explainability at all, it also paves the way to formulating concrete suggestions for a responsible use of XAI techniques that reduces the potentially negative impacts of AI systems on society. We sketch these societal implications in our concluding remarks (§5).

12:00-12:40

Anna Leuschner: From Denialism to Distraction: Reflections on Industry's New Strategies against Climate Action (Invited Talk)

Leibnizsaal

For a long time, industrial and political stakeholders have sought to postpone climate-policy mainly by manufacturing doubt over the existence of anthropogenic global warming and its impacts. However, given the overwhelming evidence of global warming during the last couple of years, more and more people have begun to understand the urgency of the situation. As a reaction, affected stakeholders are replacing the strategy of "doubt-mongering" by new strategies that seek to distract public attention. As Michael Mann has recently pointed out, two strategies are particularly salient: Blaming the Individual (BI) and Doom-Mongering (DM). I'll explore both strategies and discuss how they could be effectively addressed with recourse to Parfit's and Kernohan's arguments on "accumulative harm".

14:05-14:45

Sonja Riegler:
'In Their Best Interest' – Situated Ignorance, Expert-Involvement and Debates on Epistemic Paternalism

Leibnizsaal

In recent years, political epistemology has seen a surge of research on "epistemic paternalism". In short, "epistemic paternalism" concerns practices through which agents interfere with other agents' inquiry for their epistemic good (Jackson forthcoming: 134; Ahlstrom-Vij 2013a: 51). Often, epistemically paternalistic practices are a result of perceived "competence gaps" in agents and happen with their best interest in mind (Estlund 2021). Such "competence gaps" also play an important role in debates on "voter ignorance" and "epistocracy" (Estlund 2021; Brennan 2021). A significant amount of survey data indicates that many people are ignorant of basic political and scientific information (Brennan 2021: 146). In reaction to these findings, deliberations on "epistocracy" address the following question: If voters' competence and knowledge are very low and another group would do far better in casting an informed vote, does this mean that we should allocate more voting rights to the latter?

In my paper, I take a closer look at the potential legitimacy of epistemically paternalistic practices in politics. More specifically, I focus on the role of "ignorance" in debates on epistemic paternalism. Importantly, however, I do not endorse claims concerning the supposed need for a "rule of the wise". My aim is to debunk the myth that voter ignorance poses a threat to democracy. And yet, I concede that people's ignorance on political and other matters is a problem we must address if we want to ameliorate our political practices and citizens' political participation. In this vein, I show that the notion of "ignorance" should play a centre-stage role in debates on epistemic paternalism. Drawing on insights from feminist epistemologies, I investigate specific causes for ignorance, namely social positionality, and the epistemic resources, such as (no) access to education, that come with specific social locations of privilege or domination (see Alcoff 2007; Dotson 2012; Medina 2012; Pohlhaus 2012). I call this "situated ignorance". Deliberations on situated ignorance explain why certain people might be better informed on political and scientific matters than others. These lines of inquiry pave the way for what I label "situated rational ignorance". I believe that not everyone should be required to be informed about political or scientific matters to the same extent. In fact, the "situated-ignorance" argument I develop demonstrates that not everyone can be informed to an equal extent. For this reason, I argue that only if we acknowledge how social power structures influence people's ignorance regarding various matters can we begin to ask if certain forms of epistemic division of labor and epistemically paternalistic practices are legitimate.

Moreover, I will show that "specialist-expert involvement" in political decision making is a practice that should be endorsed and supported. I believe that it can be conceptualised as a legitimate form of epistemic paternalism that stems from a justified and productive epistemic division of labor. And yet, I argue that as soon as a group has a preconceived notion of "what is best for people", one must be very careful. How can we facilitate virtuous and inclusive expert-involvement that is responsive to the heterogeneity of people's social situatedness? To answer this question, I consider different forms of "expertise" (Evans & Collins 2009). I propose that one solution to the potential problem of domination through specialist-expert involvement could be to look at experience-based forms of expertise and expand the notion of epistemic paternalism to include these forms. My paper therefore closes with a section on the positive effects of experience-based accounts of expertise for considerations on epistemic paternalism.

Markus Ahlers:
Justification of Algorithmic Decision Making Needs not only a Good but also the Right Explanation

Karmarschraum

Philosophical debates about artificial intelligence (AI) tend to focus on high-stakes cases. They concern the problem that algorithmic decision making is opaque while large moral implications are at stake. One approach to address this problem is so-called Explainable AI (XAI) which draws on a simple idea: If algorithmic decision making is somehow explainable, we can also deal with it ethically. A good explanation is necessary to assess whether algorithmic decision making is justified. Starting from Humphrey's definition of opacity, different authors in the field of XAI look for the epistemically relevant elements of algorithmic decision making. Here, there are several approaches to clarify how opaque AI should be dealt with in its decision making. For example, the position is taken that automated decision making only needs to be reliable and that a deeper explanation is not necessary for dealing with the respective AI (Duran 2018). Likewise, however, some authors also look for causal models that can provide a mechanistic explanation of decision making (Kushner et. al. 2017). In addition, the explanation depends on the various stakeholders involved in the decision (Zednik 2018).

These positions have in common that in high-stakes cases a justification of decision making always depends on the degree of the epistemic certainty of a given explanation. In my talk I will show that the normative justification also has an influence on the epistemic relevant elements of the explanation. The social context of the particular justification determines which epistemic elements of algorithmic decision making are relevant and substantive. Different domains and social systems prioritise different normative values from which qualitatively different demands on explanations are derived. In the legal system, for example, the value of the comprehensibility of judgements has a high priority; decisions must be able to be explained individually. Statistical evidence based on group membership is not sufficient here. In contrast, in the medical field, statistical evidence may well be perfectly enough for a decision.

In this way, the degree of epistemic certainty of a given explanation is not alone sufficient for a possible justification of an algorithmic decision. Rather, the given explanation must also hit the substantive points necessary for the respective justification. And these substantive points of explanation arise from the social context of the justification. Following the example, it may well be sufficient in a medical treatment if the statistical evidence expresses a high reliability of AI, and patients are diagnosed almost 100% correctly. In the legal system, however, a different kind of explanation is needed when an algorithm is used to assess a criminal. Statistical evidence is certainly helpful here, but for the assessment, individual characteristics of the offender are equally epistemically relevant elements, regardless of group affiliations.

Here, it is important not to mistake the demands of justification for the explanation with the demands of involved stakeholders. The demands of justification arise from the social function of the respective social sector – the medical sector is supposed to provide for people's health. Stakeholder interests can represent these demands, but not constitute them. Thus, a good explanation is necessary for a normative justification of a decision, but the required justification – embedded in a social context – also determines not only the degree of epistemic certainty of an explanation, but which substantive points must be addressed in an explanation.

Hans Radder:
Do Open-access Publishing and a National Research Agenda Serve the Public Interest of Science?

Kerstingzimmer

The presentation consists of three parts. First, I summarise my conception of a public interest and explain how it may be used to analyze and evaluate specific aspects of science and technology. A public interest is an interest in realising and maintaining positive (or in preventing or removing negative) states of affairs of basic significance, that (1) affect (or will probably come to affect) either all members of the public or a specific part of it and that (2) are democratically judged to be of public import. A key aspect of this definition is a multi-dimensional concept of democracy, which should comprise voting, deliberation, a constitution, an inclusive demos and electorate, and a general right to stand for election. In the second part, I examine the possible public interest of open-access publishing (OAP). OAP has been promoted by many as a crucial value of a more democratic science. What is lacking, however, is a critical reflection on the different application contexts of this value, especially for the social and human sciences. In fact, current OAP practices suffer from several substantial problems:

- No remedy against the exceptional profits of multinational publishers;
- Hardly affordable for many researchers in the human and social sciences;
- Not affordable for scientists in poor and developing countries;
- The rise of predatory journals;
- Pressure on journal editors to accept more articles.

For these reasons, the degree of inclusiveness of open access publishing is small and its democratic support weak. Therefore, against much common wisdom, the conclusion is that this form of open access will not bring us the utopian liberation that its fervent (but naive) advocates want us to believe.

The third part of the paper reviews the recent development and content of the Dutch National Research Agenda (DNRA, or Nederlandse Wetenschapsagenda). The question is: does the DNRA satisfy the criteria for being of public interest? A critical analysis reveals the following problems:

(1) The degree of inclusiveness of the affected people is strongly constrained. In principle every citizen had the opportunity for suggesting subjects for the agenda. However, the subsequent implementation of the DNRA includes significant top-down procedures and a clear bias toward economic issues.
(2) For these reasons, the quality of its democratic support, especially its deliberative support, is strongly limited.

My conclusion from this analysis is twofold. Negatively, we have to conclude that calling the DNRA a case of citizen science (or citizen science policy) is already questionable, but it is certainly not a successful example of democratising science. Positively, this way of analysing this attempt at democratising science in detail makes explicit the relevant issues that have to be taken into account and entails concrete ways of where and how possible future attempts should and could be improved.

14:50-15:30

Liam Kofi Bright:
Du Boisian Alternative to Peer Review (Invited Talk)

Leibnizsaal

Pre-publication peer review is an expensive waste of time. Probably. In this talk I will outline the case for this proposition, reviewing the evidence suggesting that pre-publication peer review does not justify its many costs and disadvantages, as well as making a more positive - albeit somewhat theoretical - case that a better system is available. I shall argue that this alternative system is not only to be preferred on epistemic grounds, but also that it better fulfils a certain Du Boisian ideal for the proper role of science in a democratic society. There is hence both an epistemic and democratic case for the alternative put forward.

16:00-16:40

Quill Kukla:
Nonpropositional Ignorance

Leibnizsaal

The epistemology of ignorance is concerned with how ignorance is produced, maintained, and propagated by our epistemic practices. The overwhelming focus in the epistemology of ignorance literature has been on what we can call propositional ignorance, which is ignorance that a fact is true. Indeed, many accounts of ignorance in philosophy of science and epistemology restrict the notion to propositional ignorance by definition. But just as propositional knowledge is not the only kind of knowledge we can have, likewise propositional ignorance is not the only kind of ignorance we can have. Moreover, other kinds of ignorance are epistemologically and practically important. My first goal in this presentation is to develop a typology of kinds of ignorance. For example, in addition to propositional ignorance, there is practical ignorance, or the lack of a skill or inability to do something. I am ignorant of how to ride a unicycle, but my ignorance is not a matter of my not knowing true facts. There is also perceptual ignorance, or the inability to perceive or be sensitive to some type of input. My ineptness at listening to and parsing jazz is a kind of perceptual ignorance. We also have phenomenological ignorance, which is when we don't know what it's like to experience something. Frank Jackson's famous Mary the scientist, who doesn't know what colors are like, is phenomenologically but not factually ignorant. I will catalogue and briefly explore the varieties of ignorance.

My second goal is to argue that nonpropositional forms of ignorance are produced and maintained by our epistemic practices in significant and interesting ways, just as are propositional forms of ignorance. For example, the development and institutionalisation of some scientific techniques and skills goes hand in hand with the loss of or failure to develop other skills. The rise of GIS, for instance, has largely erased a large body of specialised cartographic skill. To give a quite different example, the trend in science towards multi-authored, multi-site, multi-disciplinary science means that many scientists will work at quite a far remove from parts of their own projects, and they will have phenomenological ignorance of the on the ground experience of many dimensions of the research processes in which they participate. A lot of hands-on scientific labor is offloaded to students and lab technicians, who are not the ultimate authors responsible for the knowledge produced. The resulting ignorance can hamper scientists' feel for the nuances, implications, and limitations of their own results.

My third and final goal is to argue that nonpropositional forms of ignorance are maintained by science communication, in ways that impact laypeople's ability to understand and trust science. Traditional science communication conveys propositions. But most people, even if they are well-read, are deeply ignorant of the practices of science; they lack scientific skill, perceptual capacity, and experience. The lab is a completely foreign environment to them, and scientific concepts are abstract for them. But skill and experience are part of scientific knowledge, so even good science reporting is imparting only partial scientific knowledge. In light of this, it is not surprising, I argue, that people are unsure when to trust scientists or scientific claims, or how to act in light of scientific information. Practicing scientists trust their own results on the basis of robust multimodal knowledge, which cannot be passed on through testimony. Perhaps, I suggest, we need more hands-on advanced science education and participation if we want more public trust in science. I end with a few examples of participatory science.

Samuel Maia:
Normative Validity: The Case from Poverty Measures

Karmarschraum

Here, I try to elaborate on Alexandrova's notion of normative validity (2017: 151): the property of a measure that adequately reflects its targeted concept's normative aspects (moral and political). The sense of validity with which I work here is one of the social sciences and public policy: the ability of a measure to represent a concept adequately (Jannuzzi 2005). Normative validity is a specific type of validity aimed to deal with thick concepts, i.e., concepts that express, together with descriptive judgments, normative ones (Williams 1985; Kirchin 2013; Roberts 2013). If we assume that these normative elements cannot be eliminated, we must ensure that our measurements adequately reflect them. In the field of measurement construction, this is what the notion of normative validity seeks to address.

Normative validity comes with two special requirements. First, scientists should be transparent regarding the normative choices involved and the contingency of the descriptions based on their measures. Second, it urges us to decide who should make these choices: Only experts? Or should we include non-experts? The adequate answers to these questions are contextual: they depend on what concept a measure is based on, and for what purpose it is being developed. To illustrate it, I take as an example a paradigmatic thick concept, poverty (Müller & Neuhäuser 2011; Dupré 2012; Schweiger 2020). Poverty is a valuable concept insofar as social scientists have developed several methods to make the normative aspects involved in its measures transparent. Two examples of such methods are the axiomatic approach (Sen 1976; Foster et al. 1984) and the social welfare function approach (Atkinson 1987; Deaton 1998). In both, normative validation takes place by formal means: a researcher shows that the mathematical structure of certain measures satisfies certain axioms or principles, which in turn reflect normative judgments (e.g. the Pigou-Dalton principle of transfer, according to which transferring a given amount of benefit from an agent to another who is worse off, even after the transfer, yields a less unequal distribution).

The problem of *Who chooses?* is another for which the literature has provided insightful answers. First, we need to choose how much weight should be given to non-experts: one end of the spectrum provides no room at all for non-experts (Townsend 1979 is a notable figure here), and the other gives full room (Pogge & Wisor 2016). The consensus in the literature is probably

somewhere in between (Alkire et al. 2015). If we admit a role for non-experts, we still need to choose which of them: the whole society (Mack & Lansley 2015)? Or the poor themselves (Narayan et al. 2000; Walker 2014)? It is worth noting that these positions parallel others in the values in science literature (Elliot 2017; Bright 2018; Brown 2021; Alexandrova 2022). Normative validity is only one of the different desirable properties of measures. Often, it will conflict with other possible desiderata, such as the feasibility of its operationalisation and comparability (Jannuzzi 2017). How to deal with such conflicts and trade-offs is, ultimately, a normative issue.

Torbjørn Gundersen:
Policy Recommendations, Science Advice, and Well-placed Trust

Kerstingzimmer

This paper examines the nature of public trust in scientists' policy recommendations and what makes such trust well-placed. Public policy relies on scientific evidence about a wide range of topics in the form of empirical generalisations, causal claims, and predictions. Accordingly, discussions over public trust in science have mainly revolved around epistemic trust and what it means for citizens and policymakers to have well-placed trust in the testimony of scientists. Yet, science advice does not only amount to the provision of evidence relevant to policymakers and the public but also policy recommendations about what they should do in a policy issue. Science advice institutions are often explicitly asked by way of mandate to develop, describe, evaluate, and rank possible and feasible policy proposals. Moreover, in policy issues that are technical, complex, and urgent, citizens and policymaker even rely on the scientists' policy recommendations similarly to how they rely on scientists for factual evidence. Whether the public and policymakers decide to trust such recommendations or not can have vast social, economic, and environmental consequences.

The paper examines the difference between epistemic trust and trust in scientists' policy recommendations. Matthew Bennett (2021) has recently developed a new notion of 'recommendation trust' according to which trust in expert's policy recommendation is distinct from epistemic trust: in order to be well-placed, it requires that the recommendation is based on the interests of the trust-giver, be it the public or a political body. Bennett's account amounts to a demanding standard for trustworthy expert recommendations, and, accordingly, he concludes that it is unlikely that it will ever be realised. Given the crucial role that expert recommendations play in important and technical policy areas such as environment and health I develop a less demanding account of well-placed trust in scientists' policy recommendations along two main routes that avoids the pessimistic conclusion.

First, I examine the toolbox that scientists have at their disposal in order to modify and hedge their recommendations. Scientists can make conditional recommendations based on several normative premises and they can vary the recommendation with respect to such things as specificity, scope, and normative strength. Second, I show how the distinction between basic and enhanced public epistemic trust developed by Irzik and Kurtulmus (2019) can be applied to recommendations. In this view, the public might have basic recommendation trust in a science advice mechanism to provide recommendations due to its competency, transparency, accountability, and honesty, without agreeing on the normative premises and risk assessments underlying the recommendation. In sum, the policy recommendation can be well-placed if scientists are able to formulate the recommendations properly, in a competent and responsible manner and according to reasonable expectation on the process of science advice without presupposing that the recommendation must be based on or aligned with the interests and values of the public.

16:50-17:30

Martina Vortel:
A Critical Approach to XAI: The Case of Digital Phenotyping

Leibnizsaal

The research program of XAI in medicine is motivated by the epistemic goal of understanding how predictions are made in order to mitigate the risk of intervention based on inaccurate predictions, and by broader normative goals including the prevention of the sources of algorithmic bias. However, early attempts to implement XAI in medicine – which predominantly focus on post-hoc interpretability of predictions via computational tools that probe the relative significance of various features – fall short of achieving normative goals related to justice because they do not challenge the underlying logics of AI which classify based on a colonial narrative of the human. If the research program of XAI is to take seriously the normative goal of justice, then the criteria of explanation of an AI system must include a critical assessment of the default eugenicist assumptions behind statistical modelling, the ground truth used to train the model, and the recursive dynamics of how the model's predictions are performed. To exemplify the proposed critical theory approach to explanation in AI, I examine the case of digital phenotyping (DP), a Silicon Valley-originating technology that uses smartphone user data-derived 'digital biomarkers' to predict user behavior, cognition, and affect to augment clinical decision making in the diagnosis and prognosis of mental health disorders. I argue that as long as the predictive results from a DP model are evidenced to have external validity, post-hoc interpretability of the model's internal functioning does not add much epistemic value in lieu of the goal of fairness. I argue that the program of XAI in DP should instead direct attention towards interrogating the underlying logics of algorithmic psychometrics and how DP models do not merely represent but actually construct mental health phenomena.

Drawing on recent debates in critical theory of AI, I first trace how emerging DP approaches, by using psychometric scoring of cognitive function and affect as their ground truth and machine learning-based analysis based on homophily, risk

perpetuating the eugenicist logic of the superiority of white intellectual ability. Secondly, I explore the performative nature of DP models, considering how they habituate users to act in accordance with the scaled and decontextualised aggregate, while epistemically obfuscating subjective and contextualised understanding of mental states. Such a critical theory-based explanation maintains a commitment to understanding the link between the AI system and the target system of mental health, rather than subsuming an explanation into the isolated DP model in computational terms. The resulting critical understanding of the colonial and patriarchal assumptions underlying the algorithmic psychometrics employed by DP tools can serve to empower users to better understand their epistemic opportunities and limitations, and potential harms, and hopefully encourage a commitment to a diversity of ways of modelling and approaching mental health. These findings support the call to further open up the research program of XAI in medicine toward discussions in critical theory of AI so that explanations can enable a collective critical understanding of how power and capital flow through AI, ultimately as means of intervention in associated group-level harms.

Mel Fagan:
Interdisciplinary Explanations: A Framework

Karmarschraum

Explanations in scientific practice are often highly technical and specialised - the Standard Model, the genetic code, models of viral infection, and the like. Experts in theoretical physics, molecular biology, etc., gain understanding from explanatory models constructed and used in their own field. Entering a scientific field is largely a matter of learning its explanatory models and accepting the epistemic norms that underpin them. Social aspects of explanation and understanding are a rich and relatively unexplored area for philosophical study (Woody 2015). Several recent works explore explanatory diversity across the sciences (e.g. Potochnik 2017; Rice & Rohwrer 2020; Ross 2020). This paper builds on and extends these pluralist accounts, by more fully engaging social aspects of scientific practice. Scientific understanding, conveyed by an explanatory model, is largely confined to a particular scientific community comprised of experts. If explanation and modes of understanding differ across scientific fields, how can scientific explanations be objective? How can scientific understanding transcend the boundaries of particular expert communities? Empirical studies of interdisciplinary research suggest that, often, it doesn't. Interdisciplinary research is challenging and demanding, and despite many top-down inducements its track-record and conditions for success remain murky (see e.g. Koskinen and Mäki 2016; O'Rourke et al. 2016; Frodeman et al. 2017). But sometimes interdisciplinary projects succeed. This paper proposes a conceptual framework for analysing and evaluating "interdisciplinary explanations" (IDE). Although only one strand of scientific practice (indeed of interdisciplinarity), IDE has broader implications for scientific objectivity and science-society interactions as well.

The IDE framework has three parts. The first is very general. Any IDE research project can be located with respect to two continuous axes: stage of the project when different fields interact (early or late), and degree of connection involved (low or high). Previous research on interdisciplinarity suggests two clusters of projects: early-stage high-degree connections (corresponding to transdisciplinary research), and late-stage low-degree connections. The latter are sometimes classified as "multidisciplinary research", distinct from genuinely interdisciplinary research (e.g. Leavy 2012). But these distinctions are murky and contested. The rest of the IDE framework clarifies this sprawling category of cases. Transdisciplinary research, on the other hand, is sought-after by funding agencies and policy-makers (e.g., NSF's call for "Convergence Research," NSF 19-551 2016). But it is very labor-intensive, marked by long start-up periods, continuous interaction across established specialisations, conceptual innovation or novelty, and (potentially though infrequently) emergence of a new, enduring, specialised community. Late-stage low-degree connections between explanatory models are a more promising approach for many projects. The second part of the IDE framework clarifies these connections. Very briefly: explanatory models can be connected directly or indirectly, in virtue of similarity or difference. I work through each of these four possibilities. For disparate fields, direct connections are unlikely; finding ways to connect very different explanatory models takes work. That work is a form of collaborative activity; pro-social action. Building on that idea, the third part of the IDE framework proposes norms for minimal indirect connections between different explanatory models. These connections are more substantial than those associated with multidisciplinary research, yet relatively undemanding. The basic idea is that satisfying norms of pro-social activity, for the special case of connecting distinct explanatory models, also contributes to satisfying epistemic goals for explanation – namely, understanding that transcends the boundaries of a single community of experts. In IDE, social and explanatory values coincide. The rest of the paper explores broader implications of this result.

George Barimah:
Deciding Whom to Trust and Testimonial Injustice in Science

Kerstingzimmer

Believing the testimony of experts is based on trust. Non-experts usually decide what to believe by deciding whom to believe (Anderson 2011). Moreover, due to the cognitive division of labour in science and the technicality involved in assessing scientific claims, non-experts are better off deferring to experts (Bromme & Gierth 2021; Keren 2018). Despite this, experts must earn the trust of non-experts, if their claims are to be believed. Although non-experts may lack understanding of important scientific issues, they can assess whether a scientist or an institution is trustworthy (Goldman 2001; Bromme & Gierth 2021; Anderson 2011; Keren 2018; Mercier 2020). Non-experts can assess expert trustworthiness by checking their competence, honesty, benevolence, epistemic responsibility or whether there is a scientific consensus about a claim (Goldman 2001; Anderson 2011; Levy 2019). Although non-experts, in principle, can perform the task of assessing the trustworthiness of experts, in practice, they face challenges when doing this. I propose that by sharing epistemic burdens with science journalists and scientists, non-experts can judge expert trustworthiness.

However, there are more daunting obstacles in deciding whom to trust. Some evidence from psychology shows that people usually trust those who share their values and even go further to seek out and believe claims which confirm their worldview. This phenomenon has been captured by Kahan and Braman (2006) as cultural cognition. Others also prefer to remain ignorant about the scientific consensus on an issue because they want to maintain membership of certain cherished social groups. This phenomenon has been described by Williams (2020) as 'socially motivated ignorance'. What this means is that even if an expert is trustworthy, she faces a good chance of not being trusted because she does not share the beliefs of a certain social group. In this talk I shall argue that selecting experts who confirm our beliefs and values seems rational (Rini 2017) but, in some ways, constitutes an epistemic vice which results in testimonial injustice against trustworthy experts. I share Hardwig's (1994) view that an ethics of expertise is not a one-way street, non-experts also have a role to play by being aware of their biases and acting against them when deciding which expert to trust.

17:35-18:15

Kevin Elliott:
Managing Value Judgments in Contexts of Scientific Dissent (Invited Talk)

Leibnizsaal

One of the major difficulties facing contemporary science communication and policy making is the challenge of responding to widespread misinformation and disinformation without blocking legitimate forms of scientific dissent. Philosophers of science have suggested criteria for distinguishing dissent that is epistemically productive from dissent that is epistemically detrimental, but these criteria have come under criticism. This talk suggests an alternative approach to navigating scientific dissent. Rather than attempting to identify instances of dissent that are epistemically detrimental, it suggests using the literature on value in science to develop strategies for managing the disputed value judgments that underlie most instances of scientific dissent. Using a case study involving dissenting scientific views about the treatment of Lyme disease, the talk illustrates how this approach opens up new avenues for responding to scientific dissent in a productive fashion.

Friday, 30 September 2022

09:00-09:40

Tarja Knuuttila:
Surrogate Reasoning: Representational and Artefactual Approaches (Invited Talk)

Leibnizsaal (via Zoom)

Scientific practice revolves around an amazing variety of constructed objects rendered by different representational tools and media, and enabling inferences concerning the natural and social phenomena scientists are interested in. The philosophical discussion has approached the epistemic uses of such artefacts in terms of surrogate reasoning. Insightful though this discussion has been, it has remained limited in scope in that it has tended to fuse surrogate reasoning with representation. I argue for an alternative artefactual approach that widens the discussion of surrogate reasoning beyond representation and modelling, covering various kinds of scientific constructs and the different analogical and other relations among such objects, and between them and the features of natural and social systems. I use examples from synthetic biology and economics to exemplify the artefactual approach to surrogate reasoning.

09:45-10:25

Pei-hua Huang and Samantha Copeland:
Unboxing the Black-box Problem: The Relationality of AI in Medicine

Leibnizsaal

Despite the potential for assisting in medical diagnosis, the black-box problem of AI systems developed with advanced machine learning algorithms remains a thorny issue that is yet to be resolved. Some of the major concerns focus on the epistemic side of trust-building and transparency. The lack of transparency of the reasoning process makes it difficult for medical practitioners and their patients to develop trust in the system (Price II 2015; Vayena, Haeusermann, Adjekum & Blasimme 2018). While the lack of transparency indeed could undermine one's trust in a technology, medical practitioners and patients frequently utilise technologies they don't fully understand (e.g. the image created by a fMRI). Focusing on transparency offers an incomplete understanding of the discomfort one can justifiably have with the AI systems developed with advanced machine learning algorithms.

In this paper, we look at the relational contexts of using AI in medicine, with a focus on the doctor-patient relationship when mediated by technology. We draw on literature that develops the notion of self-trust as a relational concept (Mackenzie & Stoljar 2000) and on theory of extended cognition (Clark & Chalmers 1998) to shed light on issues less addressed by other epistemological and user-oriented accounts. The cognitive assistance an AI technology is meant to play in medicine gives rise to a relationship between a medical practitioner and the technology which requires sufficient level of self-trust from the practitioner's side and trustworthiness from the technology's side. This relational approach captures well the insight that transparency alone does not guarantee trustworthiness. It is also critical that the practitioners may act on their self-trust in (1) determining which epistemic resources (e.g. technologies, guidelines, and systematic reviews) are reliable and relevant, and (2) exercising clinical and rational judgment about not only evidence and outcomes, but about the processes that lead to them. Self-trust and thereby trust in the cognitive tool are developed within relations and are constituted by the relations that develop out of regular interaction as well the socio-behavioural and professional context. We will present several suggestions for what relational aspects of not only AI systems but the context in which they are introduced and used are most relevant for ensuring AI tools are indeed trustworthy in the sense of empowering practitioners and patients.

Elis Jones:
What Role does Value Play in Socio-ecological Accounts of Science?

Karmarschraum

By recognising that science is a social activity, and situating it within its social context, much progress has been made in unpicking the multiple roles for value in scientific processes (Kincaid, Dupré & Wylie 2007; Elliott 2018). Here, I attempt to do a similar thing by treating science as a socio-ecological activity, and asking the same question: what roles do different forms of value play here? I focus on a promising area of connection between ecological theory and scientific practice: niche construction, via Joseph Rouse's (2016) 'new naturalist' account of science, whereby scientific practice is a form of niche construction performed by an organism (the scientist). I apply this to the case of coral reef science. In doing so, I draw on distinctions made in ecological literature between different types of niche (such as ecological, cognitive, fundamental, and symbiotic niches). I also include evidence from coral science literature, and interviews with coral scientists.

I look at three ways the activities of coral science might be related to niches, along with the roles for the relevant forms of value in each case. An important question to answer here is: to which entities and purposes are niches relative? First, I look at the coral scientist as an individual organism shaping its environment for its own purposes, that is, niche-construction in the

traditional sense. Some basic forms of value are relevant here: economic, biological, social, and epistemic. This level of explanation is unsatisfactory, ignoring many features of coral science (and perhaps science generally), notably that it is not conducted solely for the individual (cognitive or ecological) benefit of coral scientists. Next, I look at coral science as aimed at maintaining niches at the level of the human species. This conception finds a clear articulation in the ecosystem services programme, which is based on a conceptualisation of coral reefs as sustaining the wellbeing of humans as a whole and is common in coral science. Whilst this explains some aspects of coral science, it fails to capture many of the (particularly less anthropocentric) values alluded to by coral scientists, including in the practical use of the concept of ecosystem services. Finally, I look at coral science as a process of cross-species niche construction. Here, the relational values framework (Norton & Sanbeg 2020) is more relevant. I argue that coral science involves co-creation of a set of niches, with both ecological and cognitive dimensions, for a variety of organisms. The roles for values here are messier but more realistic, failing to accord neatly with distinctions made in various theories of value. This option best captures what is happening in coral science and offers a nuanced account of values in science from a socio-ecological perspective.

I then look at some of the implications of this view of science and values. On this account, epistemic and non-epistemic values can be seen to be overlapping and intertwined. This is linked to the overlap of cognitive and ecological niches (Bertolotti & Magnani 2017), drawing together socio-ecological niche construction theory and critiques of the epistemic/non-epistemic distinction from social accounts of science. To conclude I show that this account connects senses of value from a variety of fields, complementing existing accounts of values in science, and offering an ecologically grounded, naturalistic, and less anthropocentric picture of science and values.

Saana Jukola:
From Population-level Guidelines to Individualised Nutrition Advice? – Sociopolitical Implications of Personalised Nutrition

Kerstingzimmer

Nutrition science has traditionally relied heavily on population-level evidence, especially evidence from observational studies. However, it is facing what could be called a ‘credibility crisis’ (Penders et al. 2017; Jukola 2021). On the one hand, critics have questioned the reliability of the evidence originating from observational studies and demanded randomised controlled trials to back up claims about the link between food and health. On the other hand, the whole aim of implementing effective public health interventions and providing individual guidance based on population-level evidence has been called into question (e.g. Huovila & Saikkonen 2016; Ordovas et al. 2018). Recently, Personalised Nutrition (PN) has arisen as a challenger to the traditional population-based approach to nutritional evidence and advice. It aims at providing more effective interventions for individuals by utilising genetic, nutritional, medical, etc. information.

This talk addresses the following question: What are the sociopolitical implications of the trend towards PN? In order to provide answers, I start by drawing on Longino’s (2013) account of local epistemologies to briefly outline the epistemic landscape of PN and to ponder how it differs from that of the so-called traditional nutrition science. Despite its recent proliferation, PN still lacks a commonly agreed-upon definition (e.g., Bush et al. 2020). My hypothesis is that there are different ways of conceptualising PN and, consequently, of delineating what its central research questions and methods are. For example, there are differences in which physiological, genetic, or clinical parameters researchers focus on (e.g. Drabsch & Holzapfel 2019). Second, I analyze the potential ethical and political implications of PN. Here, I hypothesise that at least some dominant conceptualisations of PN can lead to effective public health interventions being undermined with undesirable consequences. This concerns views of PN that focus on genetic variation or epigenetic marks, while overlooking the effects of social and environmental factors, when differences between health outcomes are explained. I suggest that PN may overemphasise the responsibility of at-risk individuals for their own health to the detriment of interventions targeting social determinants of health.

10:55-11:35

Stefano Canali:
Benefits and Risks of Wearable Technology for Health: Integrating Epistemology and Ethics

Leibnizsaal

An increasing number of wearables devices, such as watches, patches, garments, are equipped with sensors that can track biomedical parameters, including heart rhythm, body temperature, respiratory rate. The use of wearables for health is presented as a promising turn towards a more precise, personal, and digital biomedical research and care. In this paper, I discuss the epistemological implications of using wearables as sources of biomedical evidence, identify epistemic issues therein, and introduce the need for an integrated epistemological and ethical analysis of wearable technology for health.

I start my analysis by discussing the epistemic contribution of wearable devices to the development and application of biomedical knowledge. Based on a review of the biomedical literature, I identify four main epistemic functions that wearables are used to serve in current biomedicine: monitoring, screening, detection, and prediction. I then argue that current uses of wearable technology to serve these functions follow the principles, goals, and methods of proposals of p-medicine, an umbrella term used to discuss the various movements aimed at increasing the personalisation, precision, person-centred features of medicine. For instance, disease monitoring based on wearables is clearly a new approach to the issue, moving away from the population-level data of epidemiology and towards the individual direction of p-medicine. In the clinical context, wearables

have been applied to intervene on e.g. physical activity and fitness, acting on individual traits and tailoring intervention to specific patients. Framing wearables as p-medicine technology is an important contribution to philosophical discussions on the reach and implications of p-medicine, as well as the understanding of the concrete role and impact of emerging health technologies such as wearables.

While the use of wearables as p-medicine technology is promising, there are significant issues to discuss. Substantial philosophical work has looked at ethical issues such as privacy and security, in connection to the 'datification' and digitalisation of health. But I argue that there is more: there are significant epistemic issues with ethical and social implications and ethical issues such as privacy have significant epistemic dimensions. I focus on two main issues in this direction, which both require an integration of epistemology and ethics. First, I look at the quality of data collected with wearable devices. Data quality is a crucial epistemic value, but is challenged by several issues in this context: the variability of wearable devices and the lack of information on the contextuality of data collection make it difficult to assess quality, raising questions on the representativity and inclusion of wearable data collection. Addressing these concerns, in turn, requires an integration of epistemology and ethics: poor or unclear data quality is a burden for different social groups and leads to unnecessary and unjust interventions. Second, I focus on an issue that is often discussed in relation to p-medicine: overestimation, which affects false positive rates at the level of screening and prediction with large datasets. Overestimation has indeed been discussed as a severe ethical issue for the application of wearables and other 'smart' devices in the medical context. But overestimation is clearly an epistemic issue at its core and, relying on the philosophical literature on p-medicine and social epistemology, I argue that attempts at solving overestimating cannot be grounded in technical solutions only and need a normative balance of ethical and epistemic priorities. I conclude by framing the paper as a contribution for our epistemological and ethical understanding of medicine and technology – identifying epistemic issues and values in the use of health technology is crucial to harness benefits and answer ethical and social concerns.

Michael Pope:
Warranted Trust and Divergent Values

Karmarschraum

Science provides crucial information for navigating the world around us, from complex policy decisions to everyday living. For many people, the adoption of scientific information is only possible if they trust scientists and science communicators to report honestly and competently. So, what warrants public trust in science? An initial answer is that science provides objective truths about the world that are untainted by personal, political, or social values. In short, the methods of science, and the consensus they can produce, are about facts, not values. But this view faces a problem: in many cases, science includes both. Scholars widely acknowledge that epistemic and nonepistemic values play an essential part in science, from the selection and abandonment of research projects to establishing thresholds for error and tuning models. Accordingly, divergence between the public's values and those that inform scientific research can undermine the trust that facilitates testimonial cooperation. In this paper, we leverage our empirical research to examine the relationship between trust in science and navigating communication about values. In two pilot studies, we measured the influence of intention and transparent communication about values on (1) participants' assessments of the truth of scientific communications and (2) trust updating about the source. We found that signaling intentions and transparency about the influence of values had a marginally statistically significant effect. As expected, willingness to accept science communication correlated with shared values. Our interest here, however, is how best to navigate cases where values diverge.

We examine approaches to values divergence based on transparency, values alignment, democratic ideals, and high disciplinary standards that could warrant trust in science while acknowledging the value-ladenness of science. After considering limitations of each view, we develop a hybrid view that emerges from two points. First, while each view is independently insufficient to address values divergence, features of the transparency and democratic views demonstrate that approaches that rely on disciplinary standards alone should be amended to address the influence and communication of values. This amendment reorients considerations of trust to what one trusts science for. Second, in view of one's aims when trusting, we turn to accounts of what renders science trustworthy, drawing on, for instance, Naomi Oreskes' argument that methodology combined with expert consensus supplies ample reasons to trust science as our best current understanding of a given phenomenon. Nonetheless, while science may be trustworthy in general because it proceeds by expert consensus and rigorous methods, the disciplinary standards view is designed to warrant trust in particular cases by demonstrating that such standards are met. In turn, we argue that the capacity for self-correction through persistence in those methods is an important, if neglected and misunderstood, feature for why trust in science is warranted – even if a particular result should turn out false upon further investigation. To conclude, we contend that our hybrid view can elucidate some of the ways in which a value-laden conception of science and science communication can contribute to ongoing debates about values.

Chris ChoGlueck:
Values as Evidence? Prospects and Limits of Assessing Values Empirically

Kerstingzimmer

Traditionally, empiricists have held values in suspicion (or outright contempt), following Hume's classic argument that normative conclusions cannot be derived from purely empirical considerations. Yet, because of their normative commitments, empiricists of a more feminist stripe have led the vanguard in collapsing this fact/value dichotomy by challenging idealised concepts of "value-free" or "value-neutral" science and articulating value-rich conceptions of scientific objectivity. More recently, in reconsidering the empirical status of values, some feminists such as Sharyn Clough and Maya Goldenberg now advocate for treating "values as evidence" by holding all value judgments to empirical scrutiny, just like what empiricists expect of descriptive claims. Surprisingly, the primary target of this "feminist radical empiricist" critique is feminist

philosopher of science Helen Longino, on account of her analysis of underdetermination of theories by evidence. In particular, Clough and Goldenberg contend that Longino treats values as normative background assumptions that are not subject to empirical confirmation. Is it true that Longino's contextual empiricism makes an exception for values? Moreover, are all values (like gender equality) really just "facts" or falsehoods as her critics contend, simply tested by scientific means?

In this talk, we defend Longino's contextual empiricism on the curbed ability of science to assess values. In contrast with the "feminist radical empiricism" approach of "values as evidence," we emphasise both the possibilities and limitations of empirical "tests" of values. Our analysis operates across levels of abstract and concrete. At the more conceptual level, we describe Longino's original program of contextual empiricism, particularly her views on background assumptions and evidential status. We contend that contextual empiricism not only allows for the empirical support/refutation of values, but Longino explicitly discusses when values can be empirically adjudicated and when not. To link this theory with practice, we use a case study from Elisabeth Lloyd's work on gender bias in research on female orgasm, especially the fruitful impact of Lloyd's work on empirical research. Through this concrete case, we demonstrate the prospects for a more tempered account than the feminist radical empiricists' on the normative power of science regarding values and evidence. Whereas Longino conceptualised how science must be done to combat hegemonic values, Lloyd's successful intervention in biological research demonstrates how such processes actually work: we can undermine the empirical status of sexist values by exposing their empirical deficiencies and provoke further empirical research to evaluate value-laden background assumptions and explore alternative theories without them. This case study itself supports and extends Longino's framework on the complex, limited, but genuine disconfirmation of values functioning as heuristics, at times, emphasising the importance of evaluating shared community standards empirically.

Our talk begins with a close reading of Longino on the category of empirical evidence. We then combat some of the specific criticisms from Clough and Goldenberg about Longino's treatment of values. Longino's critics mischaracterise her position, overstate certain claims, and rely on the very dichotomies that she aimed to collapse, such as rational empirical assessment by individuals vs. (irrational) social negotiation by collectives. Finally, we draw on Lloyd's case work to detail how sexist values like androcentrism can be empirically disconfirmed through what Longino calls "transformative criticism."

11:40-12:20

Sabina Leonelli:
The Ethics of Empiricism (Invited Talk)

Leibnizsaal

The datafication of society is said to be revolutionising how researchers investigate the world, resulting in improved scientific communications, faster data integration and analysis, and more reliable outputs. Big and Open Data exemplify the newest frontier of empiricism, and scientific success in extracting knowledge from such objects is often hailed as demonstrating the power of (increasingly automated) inductive reasoning: science as the collection and interpretation of facts about the world. Building on in-depth, long-term studies of data practices in the biological and biomedical practices, I review the multiple failures of Big Data empiricism, drawing attention especially to the intersection of moral and epistemic problems that this approach to research fails to address or even to recognise as significant, with severe implications for the reliability and the robustness of the knowledge thereby generated. The study of data practices clearly shows the crucial role of ethics within empirical inquiry, thereby calling for an alternative framing of empiricism focused on the limitations of data as research components and the unavoidable value judgements involved in using data as scientific evidence.

Organising Committee

The EES 2022 conference is organised by members of the DFG Research Training Group (*Graduiertenkolleg*) 2073 "Integrating Ethics and Epistemology of Scientific Research" (GRK2073). GRK2073 is a joint project of Leibniz University Hannover and Bielefeld University. The EES 2022 organising committee is composed of Matthias Ackermann, George Barimah, Birgit Benzing, Corey Dethier, Alkistis Elliott-Graves, Ina Gawel, Hannah Hilligardt, Anna Höhl, Hanna Metzen, David Stöllger, Morgan Thompson, Leonie Wiemeyer, and Jannik Zeiser. If you have any questions, please contact the organisers at eesconference@philos.uni-hannover.de.

The conference is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project 254954344/GRK2073/2.