CONFERENCE 'The Ethics of Trust and Expertise'

31 May to 2 June 2022 | Yerevan, American University of Armenia

ABSTRACTS

KEYNOTES

Professor Åsa Wikforss (Stockholm University) – Experts in a Democracy: Resistance and Rationality

Expert knowledge is needed to make good policy decisions, in particular decisions concerning complex societal challenges such as climate change or the Covid-19 crisis. At the same time, certain groups of voters tend to resist expert knowledge and they support politicians who ignore expertise. What explains the resistance and how can it be counteracted? To what extent is it an expression of irrationality? In the talk I examine two proposals according to which the resistance to expert knowledge is, ultimately, rational. I argue that both attempts to rationalize the resistance fail. Moreover, I suggest that the tendency of politicians and opinion leaders to politicize factual claims, blurring the distinction between facts and values, fuels the resistance. In the light of this, I argue, it is essential to keep factual claims and value claims distinct. This, also, is something the relevant experts need be aware of when engaging in public debates. In a democracy experts provide the factual input for decision making but they do not determine the political goals.

Professor Lynne Tirrell (University of Connecticut) – Truth, Trust, & Fear of Expertise

What motivates fear of expertise? Trusting experts is an act of humility, owning one's own limitations and vulnerability. Usually we are pretty good at knowing when we need help and seeking it, but sometimes denial (of our vulnerabilities, or our limits) gets in the way. Enter the predatory extremists, who market fear as a tool for profit and power. Extremists tend to mock and malign experts, ironically setting themselves up as experts on the far-fetched conspiracies they trumpet. How are ordinary people turned away from more accurate accounts of reality, and into states of fear that lead them to self-destructive behaviors? Here, an epidemiology of toxic speech can help. What we say, how we say it, when we are called upon to justify our claims, these are all patterned, governed by norms of discursive practice. Using some concepts from philosophy of language plus some (now more familiar) concepts of epidemiology, I'll

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870883. The information and opinions on this website and other communications materials are those of the authors and do not necessarily reflect the opinion of the European Commission.

address the way extremists use fear to block reason and to transfer trust away from actual experts.

Professor Paul Boghossian (New York University) – How Should We Explain Widespread Seemingly Irrational Beliefs?

At least in the United States, many people seem to believe claims that have been clearly refuted by the available evidence. For example, 35% of Americans believe that the recent presidential election was stolen by President Biden, even though there is no evidence to support that claim and there is lots of evidence against it. How should we explain such beliefs? And what role does trust, or mistrust, in experts play in these explanations?

PARALLEL SESSIONS

Melanie Altanian (University of Berne / University College Dublin) – Expert Ignorance and the Social Division of Cognitive Arrogance

"Just as there are socially designated authorities for expert knowledge in particular epistemic domains, there are also socially designated authorities for expert ignorance in particular epistemic domains", Medina (*The Epistemology of Resistance*, Oxford: OUP, 2013: 146) writes in his chapter on the "social division of cognitive laziness". In this paper, I examine this idea in relation to colonial and specifically genocidal "expert" ignorance and argue that it involves the social division of cognitive arrogance. I focus on the epistemic vice of arrogance rather than laziness, because as a vice of superiority, it is particularly pertinent to ignorance in the domain of colonialism and genocide. In such contexts, members of the dominant or perpetrator group are likely to exemplify this vice because norms of superiority are part of their group identity, which "justify" or normalize domination. Insofar as cognitive arrogance goes along with unwarranted high cognitive esteem, it arguably is more likely to prompt discrediting responses in the face of self-esteem threatening information. Such high defensive self-esteem characteristic of arrogance is particularly effective in sustaining practices of denial, hence active ignorance. This is because the motivational core of denial is a self-protective one, where individuals as well as whole groups seek to protect and defend their self-conceptions or world-views when confronted with information that creates uncomfortable dissonance. This further suggests a close relationship between cognitive arrogance and closed-mindedness. I explore these ideas in more detail based on the case of Armenian genocide denialism.

Solmu Antilla (Vrije Universiteit Amsterdam) – Testimonial injustice, expertise, and trust

A condition of non-expert trust in expertise is non-experts' evaluation that they will not be morally or epistemically mistreated by the expert if they choose to defer to the expert: the greater the non-experts' suspicion that they will be morally or epistemically wronged by experts if they choose to defer, the weaker the reason they have to trust experts. This paper characterises a theoretical obstacle for non-expert trust in expertise from the perspective of testimonial injustice (Fricker, 2007). I argue that especially when demand for expert deference and authority is high, experts are afforded an opportunity to increase their social status and relative expertise (to non-experts) by committing (intentional or unintentional) testimonial injustices toward non-experts while incurring low or no social costs to themselves. Especially if combined with an already lowered social and epistemic trust in experts, non-expert recognition of this opportunity theoretically further decreases both moral and epistemic trust in expertise. Adapting the results of Duijf (2021), I argue that epistemic expert distrust in these cases is epistemically rational if non-experts additionally recognise that the opportunity for testimonial injustice can result in epistemic loss, i.e. producing easily avoidable false or inaccurate beliefs in non-experts.

In the last section of the paper, I discuss two potential ways to avoid the obstacle of potential testimonial injustice for public trust in expertise. First, experts can develop, provide, and communicate a track record of expert trust and (relevant) non-prejudice in non-experts. In epistemic terms, this helps to ensure that experts do not discount non-experts as knowers or ascribe credibility deficits. While expert trust in non-experts is demarcated to the epistemic domain outside each particular field of expertise so as to not threaten or come at the cost of the epistemic value of expertise, I argue that credibility ascriptions to non-experts can also not generally be set to zero. Second, I argue that experts should incur costs for example from expert and public institutions if they commit testimonial injustices as experts.

Zara Anwarzai, Luke Capek, and Annalise Norling (Indiana University) – Expert testimonial failures and trust-building

Both experts and non-experts can fail in their obligations to one another. For example, non-experts might not weigh expert testimony as they should, and experts might not provide reliable information to non-experts. Our paper focuses on the obligations of experts to facilitate trustworthiness in testimonial exchanges. We look at sample cases of successful expert to non-expert testimonial interactions and failed expert to non-expert testimonial interactions. A good therapy session is our success case, and skepticism about the efficacy of masks during the COVID-19 pandemic is our failure case. We explain the failures and successes

by looking at whether, in these cases, experts met *all* obligations. We distinguish between *epistemic* and *agential* expertise as two forms of expertise that generate specific obligations. Epistemic expertise involves knowing the right information, whereas agential expertise involves knowing how to communicate that information to other agents as part of some shared goal. We argue that, in failed testimonial interactions, experts may have met some of their epistemic obligations but not their agential ones. A central example of this is when experts have accurate information but they fail to appropriately communicate that information to non-experts. We argue that agential expertise is often given too understated of a role in expert to non-expert testimonial interactions. This is because we often overlook the social dimension and corresponding obligations of expertise, and especially the extent to which agential expertise involves joint action: Expert actions succeed only if they are received by their target audience in the spirit indeed. In order for experts to facilitate trust between themselves and nonexperts— or, in order for experts to ensure that non-experts can "receive" their testimony—experts must meet their agential obligations, i.e., connect relevant information to the intentions, goals, and desires of non-experts who need that information.

Maria Baghramian (University College Dublin and PERITIA) – Trust, Science and the Question of Objectivity

Trust is essential to science because the effective conduct of science requires trust among scientists science-based policies can be implemented effectively where there is trust in the policy and the science behind it.

There is a growing consensus, at least among philosophers of science and epistemologists, that trust in science, in both above senses, has an irreducibly normative dimension. Science is infused with values both in the context of discovery and in the context of justification. Values, it has been argued repeatedly, fill the gap between theory and data and guide the decisions scientists take when engaging with risky inductive calculations. Trustworthy science then, according to this view, is not value free. Moreover, the values in question are not only epistemic but also moral or ethical.

A serious concern arising from this line of thought is whether a value laden science can ever meet the long held ideal of scientific objectivity. Will such conception of science not be prey to the dual dangers of relativism and subjectivism? Or to put it more starkly: should a value laden science be trusted? I consider the sources of these worries and briefly assess some key responses, including an influential line of thought presented by Helen Longino (1990 and 2002).

Arshak Balayan (American University of Armenia) – Permissibility of Moral Trust

Moral testimony pessimism claims that exercise of moral trust is morally undesirable or morally impermissible for mature moral agents. Two most promising strategies for vindication of moral testimony pessimism invoke ideas of autonomy and moral understanding.

According to the first, relying on moral testimony is wrong because it deprives one from autonomy and thus her actions have no moral worth. The argument from moral understanding considers an action morally worthy if the agent grasps moral reasons for the action. Acquisition of moral knowledge through testimony does not ensure that one grasps reasons for action and thus, it is argued, actions based on testimonial moral knowledge are deprived of moral worth.

After examining types of moral testimony and presenting different levels, types and usages of moral testimony, I assess each of these arguments separately. I show that acquisition of moral knowledge through testimony does not conflict with autonomy. Moreover at times autonomy needs to be balanced against other values such as doing the right thing. More importantly, I show that exercising moral trust involves exercise of self-trust and this is sufficient for autonomy. Later I show that to be acceptable, the requirement of moral understanding must be taken to be much more permissive, than it is usually assumed.

The conclusion is that exercising moral trust is morally permissible.

Aude Bandini (Université de Montréal) – Lay-expertise, lay-trust?

This presentation will address the issues of trust and expertise in a specific context: that of a sub-group of people living with type 1 diabetes (PWT1D) who developed their own system of automatic glucose management ("Loop"). Under the label "We are not waiting," they decided to outpace pharmaceutical companies and to provide PWT1D with the best possible means for treatment. Though not illegal, this initiative is not endorsed by any public health agency (e.g., the FDA or the EMA), for lack of safety and efficiency evidence. Relying on official healthcare guidelines, most physicians and nurses are accordingly very suspicious of this experimental, "do-it-yourself" treatment alternative: not only do they worry about legal liability, they also fear this system to be unreliable and potentially harmful (delivering too much or too little insulin can be lethal within a few hours).

Despite the reservations and warnings issued by experts however, the Loop initiative has gained tremendous momentum amongst PWT1Ds over the last couple years, essentially through social media. Step-by-step instructions for building one's own Loop system and then using it are delivered online for free. The algorithm used to match the insulin dosage with

blood glucose levels has been written, developed, and still constantly refined by volunteers as users report issues. The whole Loop initiative is advertised as a citizen participatory action research, not for profit, with a strong emphasis on transparency. It is made very clear that this project is highly experimental, and that participation is at each "Looper" own risks. Still, a growing number of patients (and parents of pediatric patients) are taking the plunge. But should they?

I'll argue that they should, for both epistemic and political reasons. To make that case, a couple of deeply entrenched intuitions about both expertise and trust need to be challenged. First, I will focus on the notion of "lay-expertise" which, at face value, is an oxymoron. In the context of medical humanities, patients living with long term condition are usually granted with "experiential knowledge," but there is no consensus about what this is supposed to mean. So far, this topic has never been properly addressed by epistemology. I hope that a better understanding of this kind of collective epistemic achievement, driven by values and political commitment, will shed light on a specific, though pervasive, kind of trust: trust based on self-reliance and individual empowerment (as exemplified by "do-it-yourself" procedures) as well as on one's community shared resources (including online), even against institutional figures of expertise and authority.

Federico Bina (Vita-Salute San Raffaele University) – Against domain-general moral expertise

In this paper, I claim that domain-general moral expertise is implausible. First, domain-general moral expertise is unrealistic because of scientific and technological advancement and specialization, the contextuality and complexity of moral problems, and their essential dependence on non-moral information and knowledge. Non-moral expertise always concerns relatively narrow fields: ophthalmologists and geochemists are experts in their domains, not domain-general scientific experts. The same applies to educators, psychologists, political scientists, or journalists, which we would hardly consider domain-general social experts. Likewise, moral philosophers, experts in moral reasoning and/or in its application to specific fields (such as biomedicine, business, war, AI, space), or virtuous agents can be held competent and reliable in specific domains, but not domain-general moral experts without any further specification. Second, functionalist views claim that moral experts exist by virtue of the function that they perform within human societies, i.e. satisfying moral novices' need for moral guidance (see Goldman 2018; Croce 2019). Although this need may be real, it is hard to point out who moral experts are, and potential candidates (e.g. moral philosophers, or virtuous agents) do not seem to perform this function that well. In a similar functionalist fashion, the difficulty to point out general moral experts could be explained by the fact that

domain-general moral experts have (had) in fact no specific function, big impact or utility in contemporary human societies.

In the second part of the paper, however, I show that skepticism about domain-general moral expertise does not exclude the possibility to recognize different levels of competence and reliability in specific capacities – such as moral reasoning and justification – in specific domains. Many scholars have suggested that moral expertise requires knowledge of objective moral truths. I subscribe to this thesis and argue that we should get rid of both concepts, defending a procedural view which allows us to identify reliable moral judgements, justifications, and agents avoiding commitment to controversial metaethical and normative views (Schaefer & Savulescu 2019).

Wout Bisschop (Netherlands Defence Academy) – Trust and Epistemic Courage

Trust is pertinent to our lives. We trust others, and need to trust them, because it is impossible for us to acquire particular information ourselves, because we depend on others due to our place in a particular hierarchical structure, because absolute certainty is often unattainable for us anyway, or for other reasons. But trust makes us vulnerable. Trusting someone who is not trustworthy can lead to harm, false beliefs, wrongly directed actions, broken relationships, and so on. It is important, then, to be able to determine when and whom (not) to trust. To that end, this paper develops an account of epistemic courage and uses this virtue approach to explore the normative dimensions of epistemic trust.

The idea is that to trust is to take (epistemic) risks. Risk is related to feelings of fear, and courage is the virtue concerned with handling fear, and thus risks. Some of the risks of trust are epistemic in nature, pertaining to attitudes of belief or acceptance about the truth or probability of some proposition p. One can be too afraid to trust, endorsing an epistemically unnecessary or even blameworthy scepticism. But one can also be overconfident, and naively trust whom or what one should distrust. The paper lists a number of commonly accepted characteristics of trust, examines their related (epistemic) risks, and provides an account of epistemic courage explaining what proper trust distinguishes from cowardice and naïveté in the epistemic realm.

Sara Blanco (University of Tübingen) – The *Explainability-Trust Hypothesis*: An Epistemic Analysis of its Limitations

Trustworthiness is widely quoted as a key property to enable the effective deployment of AI. However, it is not obvious how to achieve it. The literature often assumes that explanations lead to trust. This has been called the *Explainability-Trust Hypothesis* (ET). It is common to use ET to argue for Explainable AI (XAI). However, the link between trust and explanations is complex, and I argue that taking ET for granted is problematic.

The main goal of the paper is to analyse ET and point out its limitations. It has already been suggested by Kästner et al. (2021), that the nature of ET is epistemological rather than empirical. I elaborate on this issue, which leads to the conclusion that trust is contingent and dependent on the background knowledge of the truster and the quality of the explanations. These two nuances are the main epistemic limitations of ET.

ET fails due to its vagueness: about the kind of potential trusters it refers to, which kind of explanations should be offered and which kind of trust it targets. In order to tackle the last point, I propose a doxastic approach to trust. Such an approach helps to further understand the first two problems. I argue that understanding trust as a belief-state is the most fitting approach in the AI context. Specifying which kind of trust is the goal of explanations helps to understand until which extent ET is limited.

ET takes too much for granted. Because of this, it needs to be rejected and replaced for a hypothesis that overcomes its epistemic limitations. Acknowledging the limitations of ET helps to work in new hypothesis that capture better the relationship between explainability and trust. Spelling out which kind of explanations lead to which kind of trust, helps us to understand until which point it is fair to establish a connection between the two concepts.

Mark Boespflug (Fort Lewis College) & Jonathan Spelman (Ohio Northern University) – The Science of Ethics: A Defense of Moral Expertise

The fact that at least 97% of climate scientists agree that humans are causing climate change gives us powerful reason to believe it. If the same percentage of bioethicists agreed that there is a moral obligation to get vaccinated for Covid-19, would that give us comparably powerful reason to believe that there is such an obligation? We argue that it would in light of the isomorphism of ethics and science insofar as ethics parallels science in four ways. First, ethics, like science, makes progress. This is evinced in the endorsement of the rule of law, the abolition of slavery, women's rights, the environmental movement. Second, this progress has, as in the sciences, been punctuated by revolutions. Such revolutions in ethics are marked not only by paradigm replacements and subsequent radical changes in social institutions—e.g.,

laws and social norms—but also by radical professional developments in academic institutions such as the emergence of new fields (e.g., bioethics, environmental ethics), new journals (e.g., *Environmental Ethics*), conferences, and studies. Third, there is broad consensus in ethics. In addition to consensus related to areas of progress mentioned above, there are a host of further propositions that virtually all ethicists agree on (e.g., it's wrong to murder; giving to charity is good; sex-trafficking is wrong). Fourth, ethics, like science, involves considerable professional dedication. This is evinced by the fact that both ethics and science divide up labor, incentivize group collaboration (also dissent), practice blind peer-review, *et al*. Given this isomorphism, it seems that we should look not only to science but also to ethics to inform our beliefs and practices. It also suggests that just as we benefit from the creation of diverse scientific bodies like the IPCC, we could also benefit from the creation of diverse ethical bodies that would do something analogous.

Teresa Y. Branch-Smith (University of Cologne) – Moral obligation and vaccine hesitancy

The COVID-19 pandemic has brought to the forefront the hazards and helpfulness of science-based policy making in democratic societies. By involving experts in high-level advisory committees, politicians have been afforded advanced notice of how the virus is spreading locally, abroad, and projections on local health-care resources. Among the challenges of the pandemic, vaccine hesitancy —or an attitude of ambivalence regarding vaccines— has been a continuous unresolved obstacle to objectives like herd immunity. The reasons for vaccine hesitancy are complex, but many are directly connected to themes of trust, expertise and traditional sources of authority. Some of the quickest solutions have been to charge the hesitant with epistemic vices (e.g. gullibility, dogmatism). But there is ethnographic and sociological-based research that can be used to show that this is epistemically unjust and hinders attempts to make sense of vaccine hesitancy. This has forced experts, in addition to providing technical knowledge, to more rigorously consider moral questions like "is vaccine hesitancy a response to public health policy that is unduly coercive? If so, is that perception justified? And if not, under what conditions might coercion actually be justified?" Proposed solutions have included improving healthcare professional-parent communication and the recognition that evidence-based medicine alone is insufficient to inform debates about the ethical support required for lay publics and professionals. In this talk I will discuss the value tensions that surround some proposed solutions to vaccine hesitancy and explore how values are communicated and incorporated into decisions to trust based on pragmatic and moral grounds.

Laura Burkhardt (University of Bonn) – Trust in Medical Expertise – Exploring the Conceptual Links between Trust and Care

As every human being is vulnerable in some way, mutual care for each other lies at the heart of various kinds of human interaction. Medical care as a discipline is dealing with health as the very basic condition of human existence and thus especially in those contexts human vulnerabilities are revealed. Accordingly, caring for others lies at the heart of medical practice. Departing from Joan Tronto's conceptualization of care, this paper will focus on relationships of care within the medical context. I will show that forms of caring for patients merely based on reliance are not enough, but that there is one essential feature in these kinds of relations, namely trust, that is indispensable for good care in medical practice. It will be demonstrated that even though reliance and reliability play an important role within relationships of care, they are not enough to capture the whole normative dimension that is characteristic for acts of good care since they do not take into account a second-person-standpoint and are rather functional. In highlighting trust and trustworthiness as crucial elements of good care – particularly in the context of medical care – special attention needs to be paid to the interrelatedness of trust and trustworthiness on an interpersonal level and reliance and reliability on an institutional level. Thus, the structure of this paper will be two-part. Part one focusses on the conceptual understanding of trust as a relational practice. Part two consists of a conceptual analysis of the relation between trust and care. The aim then is to bring the two lines of thought together and get a better understanding of what the normative foundations of these two concepts are.

Elinor Clark (Leibniz Universität Hannover) – Tracking trust: Exploring reputation scores as a solution to misinformation on social media

Social media has encouraged wider participation in information communication, removing gate-keepers and lowering barriers to access. But this also creates new challenges for the novice public establishing who to trust, and exacerbates the problem of misinformation. Reputation scoring has been proposed as one potential response to the challenge of misinformation and fake news on social media, providing a recognised way of verifying and holding people to account for being dishonest or epistemically sloppy, and offloading the memory task from individual users onto an institution (Rini, 2017).

In this paper, I combine insights from computer science and philosophy, drawing on Bubendorfer and Chard's taxonomy (2014) to consider two designs for a reputation scoring system: 1) a centralised, global approach (CG) and, 2) a distributed, personal approach (DP).

I identify three clusters of concerns with CG, namely 1) determining the objective fact-checker, 2) isolating ideological subgroups and 3) shifting to a conservative centre, and consider how DP would address these challenges. DP allows users from all communities to trust that following the system will enable them to further their epistemic ends, and the (more) private nature of each person's reputation scores makes opportunities for reverse incentives and in-group signalling less salient.

I then explore some limitations and design challenges with DP, including, importantly, that it may simply reinforce existing epistemic bubbles. I gesture to three possible responses to this serious challenge: biting the bullet, defending the positive epistemic features of DP and proposing a hybrid account with centralised elements. I end by highlighting a number of avenues for further interdisciplinary research. Despite these concerns, I suggest that DP may be the most epistemically effective and least epistemically (and ethically) problematic way to apply a reputation scoring system, and that more philosophical attention should be given to this approach.

Samantha Copeland (TU Delft) & Pei-hua Huang (Erasmus Medical Centre) – Unboxing the Black Box Problem: the Relationality of AI and Self-Trust in Medicine

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. One of the major concerns focus on the epistemic side of trust-building and transparency. However, 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 little help in explaining 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 and on theory of extended cognition to shed light on issues less address 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 developer teams and the

authors of 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 as other behaviours.

Michel Croce (University of Genoa) & Neri Marsili (University of Barcelona) – Trusting the wrong sources: Pseudo-experts & belief in conspiracy theories

In the last decade, we have witnessed a proliferation of online misinformation that has led an increasing number of people to fall for conspiracy theories and unwarranted beliefs (Dan and Dixon 2021). Some researchers and journalists suggest that the emergence of "post-truth" sentiments is at the root of this phenomenon. Their proposed explanation is that the decline of trust in experts is key to explaining why people believe in conspiracy theories and develop uniformed views (Nichols 2017). This "*lack-of-trust explanation*", however, may be too simplistic, if not wholly misguided: at best, it only applies to some domains but not others (Kassirer, Levine, and Gaertig 2020).

In this talk, we develop a more nuanced explanation, grounded in the idea that previous attempts to explain the emergence of conspiracy theories underestimate the role played by what we call *pseudo-experts*. Relying on existing work in the epistemology (Anderson 2011; Coady 2012; Goldman 2018) and psychology (Hendriks, Kienhues, and Bromme 2015) of expertise, we identify four (clusters of) markers of expertise—*accuracy, inquiry-related goals, reputation,* and *skills*—and argue that pseudo-experts, despite reliably providing unwarranted answers to open questions in a domain, instantiate several features of each cluster.

By reviewing some cases studies (such as anti-vaxx theories and ufologist conspiracies) we illustrate how pseudo-experts actively contribute to develop and promote influential conspiracy theories, and we suggest that the perception that these figures are genuine experts is an important driver of people's acceptance of these theories.

We conclude by identifying three advantages of the proposed account: first, it shows—contra the lack-of-trust explanation—that people have not just irrationally lost their trust in experts; second, it avoids the temptation of overstating the extent to which these people are gullible; third, it provides a refined framework for distinguishing genuine experts from fake ones.

William Cullerne Bown (Independent scholar) – O'Neill's idea of trustworthiness revisited

Concern for trust in institutions and experts led Onora O'Neill to develop her idea of trustworthiness, a novel conception that parses the concept into the triple of reliability, competence and honesty. Here I develop this idea by articulating an abstract, natural conception of trust that can be found in relations between people and other entities and which covers both words (truth claims) and deeds (actions). This leads to an abstract and natural conception of trustworthiness that can be interrogated with ideas from the theory of measurement, the starting point for which is the observation that measurements (including rankings and classifications) are a common kind of truth claim. This approach is nicely suited to the institutional setting and O'Neill's emphasis on evidence as it assumes empirical assessment of substantive performance in repetitive systems governed by policies. It yields quantitative interpretations of reliability and competence in which they address two distinct forms of defect that may afflict supplies of truth claims or actions. A fourth element is added to O'Neill's triple, certainty, that is again drawn from the theory of measurement and concerned with defective supplies. Honesty stands apart as an aspect of trustworthiness that lies beyond the theory of measurement, becoming important exactly when the other lacks measurements. It can only be present in higher organisms that have a theory of mind, and their creations, including institutions. I come to see it as a special kind of care for an other (or others), which explains why we usually do not find a white lie troubling. I show that the idea of honesty found in studies of animal and social signalling can be derived from my idea of trustworthiness by making a simplifying assumption. I then return to O'Neill's original concerns and consider the ethical implications for institutions and experts, including the law.

Agnes Díaz Castellano (Università degli Studi di Genova) – Technology and Expertise: a redefinition of the concept of expert through collective intelligence

Expertise and knowledge are intuitively seen as a source of trust (Fletcher, 2009). For this reason, since the 1970s most governments have relied on experts for creating better institutions, intending to gain legitimacy among citizens (Warren, 2018). Because of the distribution of knowledge in society an expert is understood as a person who has a recognized authority in an area of science. But the scientification of law and policymaking comes with ethical dilemmas that lead to a lack of trust and arguably, a threat to democratic values. The tendency to scientification collides with the growing tendency to incorporate societal voices in the policymaking process, to create a more open, inclusive and participatory system to reinforce democratic principles.

Nonetheless, these debates oversee an essential element, the radical changes produced by technological advances. Therefore, I believe it is important to reevaluate the concept of expertise by taking into account the technological possibilities of the current era. In this line, I will defend the division of the idea of "expert" into three groups, the first will be based on the source of the expertise, the second on the temporal relevance of their input and the last, on its individual or collective form. I will focus on the different individual forms of expertise to try to define their ethical and democratic limitations. I will try to reconstruct parallelisms between collective forms of expertise and two forms of citizen duties, the duty to vote and the duty to become part of a jury. On one hand, the voting example will allow us to understand some of the ethical and normative values we ascribe to citizen participation. And on the other hand, the jury court will work as an example of collective intelligence as well as of the combination of different forms of expertise.

My claim is that, since there are limitations in every type of expertise, an epistemic democracy cannot rely only on one form of "expert" at a time since it generates some relevant democratic and ethical problems. Therefore, I will claim that there is a necessity to incorporate societal voices as forms of expertise in a collaborative way combined with traditional forms of expertise. By using the possibilities technology provides us, collective intelligence can be a new form of expertise that is capable of avoiding ethical and democratic weaknesses.

Steven Donatelle (American University of Armenia) – The Place of Values in Science and Expertise: Contrasting Goldenberg and Douglas with Wikforss and Kitcher

I will approach the issue of public trust in science and experts through the latest works of Goldenberg and Douglas. Goldenberg (2021) states that:

While the arguments over vaccines are often centered on the science… science largely serves as a placeholder for the values at stake…. The evidence… serves as proxies for the values that are on the line… these [value] issues are [not] easily settled and, importantly, none will be settled by the science... (p. 14)

The book argues against the idea that there is a War on Science and that the real conflict is over values.

The place of values in science and society are the focus of Douglas's most recent book where she states that:

… the understanding of science we need the public to have includes not just the role of evidence but the role of values in scientific practice…. that values can legitimately influence

the acceptance and rejection of scientific claims—and this applies to both scientist and the nonexpert public. (p 123-4)

Douglas proposes an approach in which the responsibility for building trust in science and experts is equally shared by the public and experts. This requires a new understanding of science in which values are allowed a place both in science and in the public's evaluation and acceptance of science.

I hope to be able to contrast their approach with the approaches taken by Asa and Phillip Kitcher. Wikforss focuses on knowledge resistance and motivated cognition. I will argue that motivated cognition often takes place when 'science serves as a placeholder for values at stake.'

I will contrast the place that Douglas and Goldenberg give to values with what Kitcher does in his book Science in a Democratic Society where he lays out a quite different place for values in science and public debate.

Domingos Faria (University of Lisbon) – Disagreement in Trust

We need to trust other people and groups, but there is often disagreement about whom to trust. In this talk we want to address the following question: what is the rational response in the face of disagreement over whom to trust? Namely, what should we do when we are aware that others do not trust those we trust? In such cases, should we diminish the degree to which we trust? Or should we rather ignore the distrust of others and remain unwavering in our trust?

In order to answer these questions, we will extend the theoretical framework we developed for the epistemology of disagreement in general (cf. Faria (2022)) to accommodate the particular cases of disagreement in trust. The main idea of our theoretical framework for disagreement in general is to argue for a gnostic prescriptive norm of disagreement that is based on a knowledge-first epistemology (cf. Williamson (2000); Lasonen-Aarnio (2021); Littlejohn and Dutant (2021)). According to that norm, in cases of disagreement about whether p, one must: hold steadfast p if and only if one has good cognitive dispositions in believing that p (that is, in forming or retaining the belief p, one exhibits dispositions that tend to manifest epistemic quality states – knowledge – in normal counterfactual cases).

Our aim is to expand our theoretical framework to disagreement in trust. To achieve this aim, we will introduce the notion of "trustworthy" defined as follows: S is trustworthy with regard to ϕ if and only if S has good cognitive dispositions of ϕ-ing. On that basis we can have the following prescriptive norm of disagreement in trust: in cases of disagreement about whether

to trust S's ϕ-ing, one must hold steadfast to trust S's ϕ-ing if and only if one has good cognitive dispositions in believing that S is trustworthy with regard to ϕ.

Mirko Farina (Innopolis University), Artur Karimov (Innopolis University) & Andrea Lavazza (Centro Universitario Internazionale) - Echo Chambers and the Ethics of Communication During the Covid-19 Pandemic

In the first part of the paper, we discuss one the main problems characterizing the spread of fake news during the COVID-19 pandemic, which is echo chambers. Echo chambers are social and epistemic structures or environments in which opinions, leanings, or beliefs about certain topics are amplified and reinforced due to repeated interactions with a close system; that is, with a rather homogenous sample of sources or people, which all share the same tendencies or attitudes towards the topics in question (Nguyen, 2020). Echo chambers are particularly dangerous phenomena because they prevent the critical assessment of sources and contents, thus leading the people living within them to deliberately ignore or exclude opposing views. In the second part of this paper, building and expanding on previous theoretical and empirical work, we argue that the reason for the appearance of echo chambers lies in the adoption of 'epistemic vices' (Cassam, 2019). We examine which vices might be responsible for the emergence of echo chambers and -in doing so- we focus on a specific one, which -following Dotson (2011)- we call 'epistemic violence'. In assessing and evaluating the role of this epistemic vice, we note that it can be triggered by epistemic contexts characterized by high stakes (such as the COVID-19 pandemic) that may turn ordinary intellectual virtues (such as skepticism) into vices (such as denialism). In the third part of this contribution, we suggest a way for dealing with echo chambers. The solution focuses on advocating a responsabilist pedagogy of virtues and vices (Pritchard, 2013; Zagzebski, 2018) whose main goal is to promote the establishment of mutually beneficial epistemic and liberal attitudes. We note that the notion of noetic rights can be used a positive addition to the building of a responsible epistemic environment. Doing so, we conclude, will ensure ethical communications, and prevent the future emergence of echo chambers.

Yuhan Fu (University of Sheffield) - Can We Trust AI's Moral Judgements?

In my talk, I will focus on moral testimony concerning AI: do AI have morality? To what extent do we trust moral judgements from AI? And do we even want an AI moral authority? I argue that AI (probably) does not have moral beliefs, hence it fails to have a whole grasp of moral understanding. We can trust AI's moral judgements, but we cannot treat AI as moral authority.

To illustrate and justify my claim, I will introduce an AI called Delphi released by Jiang and colleagues (2021). Delphi is an AI designed to make moral judgements. Its programme is based on what AI researchers called a deep neural network (Goodfellow, Bengio, and Courville 2016), which is a mathematical system which attempts to mimic the web of neurons in the brain. This neural network attempts to learn moral norms from 1.7 million of everyday human ethical judgements made by people in the US. Through learning and training, Delphi can answer three different moral tasks: free forms (kill a bear to save your child), yes and no questions (should we welcome refugees?) and makes moral judgements in moral dilemmas. Delphi has demonstrated 92.1% accuracy compared to human moral judgements. This has been taken to show that Delphi understands moral concepts and makes moral judgements in complicated moral contexts.

I argue that from the current performance of Delphi, AI can make moral judgements on their own. However, we cannot trust AI's moral judgements because current AI does not possess moral understanding: in order for us to trust others' moral testimony, the moral experts should understand moral facts, which current AI lacks.

Seth Goldwasser (University of Pittsburgh) – The Cure Is Worse Than the Disease: On the Concepts of Health and Disability

What difference is there, if any, between cautioning pregnant people against drinking alcohol and cautioning prospective parents against selecting potential offspring that will develop a disability like deafness or Down syndrome? According to some self-avowed eugenicists, there's none: it's permissible to select or treat potential offspring on the basis of genetic counseling or through genetic enhancement, provided doing so doesn't decrease the offspring's chance of a good life or interfere with the wellbeing of others (Veit et al. 2021). Genetic counseling and genetic enhancement, they claim, are medical tools to be used towards improving wellbeing. This paper argues that approval of genetic counseling or gene enhancement in the case of disability begs the question. Specifically, approval of genetic counseling or genetic enhancement in such cases requires a view of what types of states are disabling, i.e., which states constitute real limits on or decreases in human function or ability. But accounts of human function and health that eugenicists may appeal to underdetermine, for several disabilities, whether those disabilities limit or decrease human function or ability. Indeed, it might be that the relevant disabilities are simply other ways of being human. Moreover, we fail to see that the relevant disabilities might constitute other ways of being human because the concepts of health and pathology we inherit from ableist traditions in medicine fail to be morally neutral towards disabled people. This failure of recognition harms disabled individuals through the formation and application of normative judgments of medical professionals against selecting for disability. The very expertise that is supposed to heal in these cases very

often harms. If this is right, then there are morally relevant differences between some uses of genetic counseling or genetic enhancement and others. These differences block approval of genetic counseling or genetic enhancement in the case of disability.

Suddha Satwa Guha Roy (University of Manchester) – Trust and Reactive Attitudes

Following Strawson's seminal essay 'Freedom and Resentment' (1974) there has been much discussion on reactive attitudes. They can be objective as well as participant attitudes. Participant reactive attitudes could be personal, impersonal and self-reactive attitudes. The concept of participant reactive attitudes has enjoyed a central place in the discussions of trust, courtesy Richard Holton (1994). Holton claimed that when we trust we have a readiness to take particular reactive attitudes towards the trustee – resentment should trust be betrayed, gratitude if trust is respected. This readiness to evoke the relevant reactive attitude, according to Holton, shows that when we trust we take a trust stance – a participant stance (which Strawson called participant attitude) – towards the trustee. This claim has been significant to many subsequent philosophical engagements with trust (Jones 2004; McGeer 2008; Hawley 2019).

This paper extends Holton's thesis of trust and reactive attitudes. I argue that there are different kinds of interpersonal trust – personal and impersonal trust. One major difference between the two is in the kind of reactive attitudes they involve: personal reactive attitudes, I argue, are appropriate for personal trust whereas impersonal trust involve impersonal reactive attitudes.

Strawson claimed that personal reactive attitude is relevant when one participates personally, involving one's own interests; and impersonal reactive attitude is relevant when one participates impersonally, involving interests 'not simply' one's own. I will offer a reading of personal and impersonal participation – what counts as interests of one's own and 'not simply' one's own – different from the one given by Strawson. This, I believe, will bring out interesting ways in which we can understand reactive attitudes – both in personal and impersonal trust. Although Strawson also discussed the issue of self-reactive attitudes but I will not discuss self-trust and thus a discussion of self-reactive attitudes is beyond the scope of this paper.

Joshua Hobbs & Andrew Kirton (University of Leeds) – Deference, Trust and the Expertise of Lived Experience

Theorists of solidarity argue that activists (and others seeking to remedy injustice) ought to defer to the lived experience of those facing injustice. This call to attend to the lived experience of individuals and groups facing injustice and oppression chimes with the popular mood – as evidenced in the activist refrain to 'educate yourself'. On this account, lived experience of oppression is viewed as epistemically valuable, functioning as a certain sort of expertise.

We argue that this epistemic focus, although important, misses valuable aspects of the activist practice of deference to lived experience of injustice, which are highlighted when this is viewed through an analysis of the practice of trusting (or indeed deferring to) experts in other walks of life.

Trusting experts is not simply a matter of drawing information from them, but of relating to them, and valuing them in a certain way. We argue that understanding deferring to lived experience of injustice as a version of the practice of trusting experts highlights (at least) two normatively valuable aspects of the practice:

i. Insofar as trusting others renders us vulnerable to them, this practice can function to mitigate the power differentials that characterise the relationship between individuals facing injustice and their would-be allies.

ii. As trusting others is to value them, this practice serves an important normative function, recognising the intrinsic value of marginalised groups and individuals facing injustice and oppression.

Alex Horne (University of Cambridge) – Epistemic angst is social problem

Most of us do not enjoy admitting we were wrong. This has public and private dimensions. Publicly, it is embarrassing. It also leads others to doubt we are reliable knowers. Privately, we worry they are right to do so. Angst concerning our good epistemic standing in our community therefore goes to the core of how we perceive ourselves: as competent agents capable of making sensible decisions about how to live our lives. Consequently, many of us have a legitimate though often misfiring fear of being "outsmarted" or misled – and having this exposed – whether by our epistemic peers or experts. A symptom of epistemic angst is domain-specific confirmation bias: discounting expert or peer evidence that contradicts our firmly held beliefs about some portion of the world. Ironically, doubt concerning our own epistemic capacities therefore partially explains our doubt concerning the epistemic capacities of others.

So, while Socratic self-doubt can be a virtue, this paper concerns doubt in its manifestation as a social *vice*, as an interpersonal cartesian doubt corroding epistemic trust between ordinary citizens, and between those citizens and experts on matters of great social, political and scientific import.

One apparently attractive strategy for alleviating the social problem of epistemic angst is as follows. We should promote the idea that there is nothing wrong with being wrong. That is, being wrong does not discredit you as a knower or agent: *it happens to the best of us*. There should be no shame. You were not "beaten" by those holding the opposing view. Ascertaining the facts is not a competition for knowledge or acclaim. It is – for cognitively limited creatures like ourselves – a joint, inevitably error-ridden and deeply imperfect enterprise.

One tempting realizer of that strategy is to direct attention to the fallibility of experts: for if even they very often get things wrong, then surely there is no shame in the rest of us getting things wrong just as often. But the catch is that promoting that very idea can lead to the caustic interpersonal doubt against which promoting the idea of widespread fallibility was supposed to provide a bulwark. First, because it can lead to doubt concerning expertise. Second, because it threatens our faith in the good epistemic standing of our community. Third, because it devalues truth's already deflated currency. If there's nothing wrong with being wrong, then why care so much about getting things right? The paper investigates whether this strategy nevertheless retains its appeal and, if it does, how best to mitigate its unwanted side-effects. It concludes by considering two further problems that arise when we attempt to do so.

Gayane Hovakimyan (Armenian State Pedagogical University) - The ethical obligations of experts advice

In modern societies, mass media, information, political discourse invades the public space, mostly all public platforms and leaves a small place for experts to raise their voices, to ensure that their expertise has a key role in any decision making process.

In the growing tendency of decline of expert's role in public life, experts community has a risk of becoming a separate self sustaining system, where they have less channels to advocate for their roles. Especially in the countries, where research and evidence based practices are not well established, and these channels are not well connected to theory and practice, through academia to reality, the risk of diminishing of experts is real and sometimes it is an argument for justifying non effective decisions.

What is the role of experts? Do they have an ethical obligation to ensure that the expert advice is used and taken into account? How can the experts advocate their professional interests and how they can merge the concept of expertise to the concept of public interest? Is it possible for experts to be neutral and to work in the specified field, to create product and also to promote the value that they bring to the field? Sometimes it is very difficult for experts to

create the value, to work together to establish the platforms where this advice is introduced to larger public and applied by politicians and Government Authorities.

Surely, there is another dimension, that expert advice is only for professionals and policy makers, and larger public does not need to be aware and to trust them. However within the growing change and more alienated relations between the real democracy and its adjustment in the world, the majority may have an important role in decision making without having any idea of the content of the subject. And if the majority is not well aware of experts, does not have trust towards expertise, there is a risk that larger public will bring and support decisions in the political arena manipulated by different non professional actors.

While it is more difficult to set up normative regulations for ensuring obligations to experts for their advice, it is key to establish ethical obligation or include this obligation into professional ethics. By this the experts will be more responsible not only for the creation of value and advice, but also for sharing this advice, disseminating and educating the public, for taking the leadership in their roles of making this trust through dialogue, open discussions, platforms and easing access to experts' advice for larger public.

So the proposed abstract looks for the advanced practices, solutions, remedies and formats where this dialogue is established and functioning effectively, also looking for the professional ethics practices in various countries and suggesting the advanced and efficient ways of creating the scope of expert ethical obligation.

Silvia Ivani (University College Dublin and PERITIA) & Alfred Archer (Tilburg University) – Science, Admiration, and Respect

Public engagement (PE) is one of the fundamental pillars of the European programme for research and innovation *Horizon 2020*. PE practices aim at bringing citizens to the forefront of scientific decisions by encouraging collaboration between scientific experts and lay people in several stages of the scientific and technological process. Promoting dialogical practices would make it possible to access citizens' valuable epistemic and non-epistemic resources (e.g., local knowledge and moral values), which, if integrated in scientific decisions, would help addressing societal challenges and achieving epistemic progress.

However, the collaboration between experts and the lay public does not always go smoothly. Recent surveys reveal that scientists often see citizens' contributions as trivial or unrealistic and that they fear that these contributions may significantly impair creativity and freedom in science (e.g., Carrier & Gartzlaff 2020). Moreover, studies report that citizens sometimes feel silenced and disrespected when engaging with experts, e.g., when their contributions are ignored or easily dismissed (e.g., Goldenberg 2021).

This paper analyses the role of respect in PE by investigating whether feeling respected/disrespected may play a role in decisions to engage with and trust experts. Moreover, it suggests that some forms of respect (e.g., recognition respect) may provide us with a better base than other forms of respect (e.g., appraisal respect) for developing and engaging in epistemically fruitful and ethically sound PE practices.

Antti Kauppinen (University of Helsinki) – Echo Chambers, Emotions, and Distrust

It is possible for people to disagree about evaluative truths while agreeing about non-evaluative facts. But as debates about political polarization have shown, people with different value orientations seem to share less and less common ground on many factual issues. It is not just that American conservatives and liberals disagree about the relative priority of climate action and short-term prosperity, but also on facts about human contribution to climate change.

While some such factual disagreements between what we might call political tribes can be explained in terms of 'epistemic bubbles' – the sheer unavailability of relevant information due to e.g. algorithmic filtering on social media– many of them persist even if information is accessible. Rather, on many issues, there are what Thi Nguyen calls 'echo chambers', social structures that "exclude by manipulating trust and credence" (2020, 142). Given the putative grounds they have for distrusting ideology-incongruent information, members of an echo chamber may be subjectively rational in rejecting what is supported by objectively best evidence.

While there is a growing literature on the political and epistemic significance of echo chambers, the role of *emotions* in their formation and dissolution is relatively underexplored. I argue that echo chambers exploit emotional identification with one's in-group to discredit outsiders' testimony and arguments. They can be formed *spontaneously*, by way of social reinforcement of motivated reasoning (such as confirmation bias) among people who share a value orientation, as well as *deliberately*, by way of conscious manipulation that also tends to exploit emotional triggers. Spontaneous echo chambers are harder even for a conscientious inquirer to spot, since there is no evil mastermind – we might be inside one right now.

Given the emotional roots of such socially motivated epistemic distrust, the best antidotes for it are emotional as well. For individuals on the inside, they include adopting a stance of epistemic humility and engaging cognitive and affective empathy that is regulated by an aspiration to impartiality. Because such remedies are available to nearly all, people bear more responsibility for remaining trapped in echo chambers than philosophers like Nguyen (2020) allow for. From the outside, effective interventions must appeal to the values that members

emotionally identify with, like the Ukrainian President Volodymyr Zelenskyy did in his speech to ordinary Russians trapped in a deliberately constructed echo chamber at the beginning of the Russian invasion of Ukraine in February 2022. Such appeals can be so powerful that those who wish to maintain an echo chamber may only be able to do so by turning it into an epistemic bubble, that is, simply shutting out information that they can't discredit. But bubbles are much more easily burst than chambers.

Nikolas Kirby (Harvard Kennedy School) – Distrust, Polarisation and Disinformation

Recent revelations delivered by former Facebook executive Frances Haugen to U. S. Congress, confirmed what researchers have long suggested: the activities of members on social media platforms in the aggregate tend to increase affective polarisation within populations, and that such polarisation also feeds back into social media usage driving more activity by members. Thus, social media companies have an incentive to increase polarisation, and indeed appear, actively, to do so.

We might think this is not merely a bad state of affairs, but also wrongful. It must be wrongful to cause and knowingly profit from causing affective polarisation, or at least to actively intend to do so. However, what exactly is wrongful about increasing affective polarisation, intentionally or at least knowingly?

The debate in large part has focussed upon 'disinformation', with the apparent assumption that polarisation is wrongful when and because it is caused by disinformation. Yet, I suggest that the act of spreading disinformation fails to explain what is wrongful about creating affective polarisation. First, much of such polarisation is caused by the selection and amplification of unrepresentative truths not falsities. Second, disinformation is typically thought to be wrongful because it deceives and manipulates, yet polarisation also harms those who must live with the deceived and manipulated.

My argument, instead, will be that we should understand the wrong of intentionally creating polarisation, as a wrong of intentionally creating *distrust* between others. Such distrust is wrongful because it intentionally degrades the quality of each person's moral position. In short, we can only justify our resentment against organisations like Facebook, and indeed other unjustifiably 'polarising' actors, by accepting that such agents, perhaps all agents, have a duty to not intentionally create distrust between others. I shall then explore the broader implications of this claim.

Joshua Seth Kleinfeld (Northwestern University) – Social Trust in Criminal Justice

What is the metric by which to measure a well-functioning criminal justice system? If a modern state is going to measure performance by counting something—and a modern state will always count *something*—what, in the criminal justice context, should it count? Remarkably, there is at present no widely accepted metric of success or failure in criminal justice. Those there are—like arrest rates, conviction rates, and crime rates—are deeply flawed. And the search for a better metric is complicated by the cacophony of different goals that theorists, policymakers, and the public bring to the criminal justice system, including crime control, racial justice, retributive justice, and social solidarity.

This Article proposes a metric based on the concept of social trust. The measure of a well- or poorly functioning criminal system is its marginal effects on (1) the level of trust a polity's members have toward the institutions, officials, laws, and actions that comprise the criminal justice system; (2) the level of trust a polity's members have, in virtue of the criminal system's operations, toward government generally (beyond the criminal justice system); and (3) the level of trust a polity's members have toward one another following incidents of crime and responses to crime. Social trust, we argue, both speaks to an issue at the philosophical core of crime and punishment and serves as a locus of agreement among the many goals people bring to the criminal justice system. The concept can thus be a site of overlapping consensus, performing the vital function of enabling liberal societies to make policy despite disagreement about first principles.

Carline Klijnman (University of Genoa / University College Dublin) – Deliberative Epistemic Democracy and Public Credibility Dysfunction

Deliberative epistemic democrats hold that democratic decision-making is valuable due to its epistemic merits, in virtue of the egalitarian features of public deliberation. This epistemic value of democracy is typically understood instrumentally, as approximation of a procedure-independent standard of correctness or goodness. Over the last few decades, the way citizens obtain and consume information has changed drastically, the most prominent developments being the introduction of the internet and social media. This paper aims to give an analysis of how these developments have impacted the workings of public deliberation by employing analytical tools from social epistemology in general and the epistemology of testimony in particular. I argue that these developments thwart the possibility for epistemically healthy functioning public deliberation, in as far as they affect crucial mechanisms of testimonial trust.

As a shared social epistemic practice, the functioning of political deliberation is largely dependent on the effective exchange and uptake of (expert-) testimony. In order to gain justified beliefs from (reliable) testimony, citizens need to be able to recognize good reasons

for trusting someone's say-so and/or be able to detect reasons for doubting the credibility of sources. This ability to discriminate between reliable and unreliable testimony is in turn influenced by one's wider epistemic community.

In the current (online) epistemic environment, the conditions for successful knowledge transmission through testimony are under threat, risking what I call *public credibility dysfunction*. Briefly put, this identifies a state wherein citizens have become uncertain about which information sources to trust, or worse, end up trusting unreliable sources. Public credibility dysfunction poses significant risks to instrumental democratic legitimacy: it preserves ignorance, increases false beliefs and even affects the modal profile of our true beliefs. Additionally, in as far as it affects citizens' standing as testifiers and hearers in deliberation, it hints at potential implications for procedural legitimacy.

George Kwasi Barimah (Leibniz University Hannover) – Epistemic Trust in Scientists: A Moral Dimension

The epistemic dimension of trust in science has received some attention in the philosophy of science, especially concerning how trusting a source for knowledge can serve the epistemic goals of collaborating scientists (Wilholt 2013; Rolin 2017), and the epistemic ends of non-experts (Irzik & Kurtulmus 2019; Anderson 2011). However, articulating the moral dimension of epistemic trust in science, with particular focus on the relationship between experts and non-experts, would benefit from more attention. I intend to do this by examining the nature of affective trust in scientific experts and how that grounds experts' epistemic and moral responsibilities to non-experts.

This talk is divided into four sections. The first section shall spell out the epistemic dimension of trust in science. The second section shall develop an account of the moral dimension of epistemic trust in science by building on the work of Paul Faulkner (2007) and Philip Nickel (2007). In the third section, I apply the philosophical analysis in the second section to the case of public expert testimony. I shall then argue, in the fourth section, that the norms of sincerity, honesty and transparency are essential for ethical science communication.

Thirza Lagewaard (Vrije Universiteit Amsterdam) – Moral encroachment and disagreement with non-dominant groups

There is a growing literature on *moral encroachment*: the idea that the moral features of a belief influence the epistemic status of that belief. In this paper the moral encroachment thesis is applied to disagreements between a dominant and non-dominant group.

In such disagreements, due to epistemic injustices, there can be a lack of trust in the expert testimony of non-dominant people.
People from non-dominant groups can have privileged knowledge about their own situation. In some instances, their testimony on their own situation could be considered expert testimony. Due to epistemic injustices, the expert testimony of non-dominant people is not always considered expert-testimony or even as valid evidence at all.
Suppose person A from a dominant group disagrees with person B, who belongs to a group that is prone to be epistemically oppressed. Suppose further that the disagreement concerns and issue that person B is in a better position to know about. For example, a man disagreeing with a woman about an instance of sexual harassment or a white person disagreeing with a person of color about a specific occurrence of racism. Such a disagreement provides Person A with a good reason to at least reconsider their belief.
The question I focus on in this paper is whether in such a case *moral encroachment* plays a role in the justification of A's belief.
I consider whether moral encroachment would apply for person A. Then, I consider an argument against moral encroachment in the context of epistemic injustice.
This paper aims to further investigate the plausibility of the moral encroachment thesis. At the same time this paper aims to delve deeper in epistemic injustice-based disagreement.

Jonathan Matheson (University of North Florida) – Does Expertise Threaten Autonomy?

There's an apparent tension between expertise and epistemic autonomy. For almost anything you may want to think about, there is someone who is in a better epistemic position than you are to determine the answer to your question. This fact seems to make thinking for yourself a recognizably less reliable route to the truth, and thus not a recommended course of action. So, our access to expert opinion seems to threaten the value of autonomous thinking.

The resolution to this tension comes in two steps. First, epistemic autonomy must be distinguished from intellectual individualism. While there is a temptation to equate the two, a proper understanding of autonomy reveals that it is compatible with intellectual interdependence and a deep reliance on others. I argue for a relational conception of epistemic autonomy which does not conflict with an intellectual division of labor and a heavy reliance on expert opinion.

Second, I stress the importance of coupling epistemic autonomy with intellectual humility. When autonomous thinkers are intellectually humble, they can take on intellectual projects without the danger of being led away from expert opinion. Intellectually humble thinkers own their intellectual limitations and are thus not in danger of sticking with their own conclusions

when their conclusions conflict with the received expert opinion on the matter. Thus, humble autonomous thinkers can do their own research, and think for themselves, without ignoring expert opinion or sticking with conclusions that conflict with expert opinion.

Thomas Mitchell (University of St Andrews) – Evidence and the Reasons of Trust

What kinds of reasons count in favour of trusting someone? This question is often framed in terms of two types of reason: practical and evidential. Among purported practical reasons are good relations with another and the benefits of efficient cooperation. These reasons are dismissed by some philosophers, such as Hieronymi (2008) and Marušić (2015), as being about the attitude itself, rather than its content. Just as it is not rational to believe something because one is offered a reward for doing so, we should not trust for practical reasons. The right kinds of reasons are *object-given*, but practical reasons are thought to be *state-given*.

So, perhaps the only genuine reasons to trust are evidential. But this is problematic. Without good evidence, trust will be irrational. But with good evidence, trust seems to be precluded; the evidence would lead us to form an ordinary rational belief, with nothing distinctively trusting about it. Furthermore, as Jones (2012) points out, cooperative relations are part of the purpose of trust, so we should not lightly dismiss them as reasons.

These considerations suggest that the usual framing is too simple; for each type of reason, practical and evidential, some are applicable and some are not. Practical reasons count insofar as they are object-given. Evidential reasons must allow for the distinctiveness of trust. This paper provides a theory that satisfies these restrictions. It will be shown that, taking a certain plausible view of trust, we can non-arbitrarily include what are intuitively the right kinds of reasons, both evidential and practical, while excluding others. This view of trust is *reliance on another's trustworthiness*. The content and type of attitude require evidence to be distinctive of trust and permit of object-given practical reasons. The result is a simple and intuitive theory of the reasons that justify trust.

Attila Mráz (Sciences Po Paris CEVIPOF / ELTE Institute of Philosophy) – The Democratic Value of Electoral Competition for Trust

This paper explores the implications of an account of the right to stand for election as a right to compete for voters' trust (Lever & Mráz, 2022). What kind of trust should candidates compete for in a democratic election? In this paper, I argue that a cognitive account of trust is

adequate to this account of the right to stand for election because it realizes the democratic value of electoral competition for trust.

First, I show that an affective theory of trust is not suitable to realize the values that ground the right to compete for voters' trust, and a cognitive account of trust is necessary to assume instead. On Lever & Mráz's account, such right is grounded in democratic equality. Members of the political community should have a right to compete not merely to win any affective attitudes of others towards them. True, ample opportunities to gain others' trust understood as an affective attitude may have instrumental value for democracy—e.g., by countering social segregation. But the right to compete for trust can only serve to realize the equal status of individuals in the political community if it involves competition for a judgment of these individuals' trustworthiness. Second, I argue that a cognitive account of trust is sufficient to ground the right to compete for voters' trust. Equal status in the democratic community requires a fair opportunity to be evaluated based on one's trustworthiness.

Finally, I address an objection. Namely, that competing for trust is an overly idealistic account of democratic elections, as voters often cast their ballot based on distrust of some candidates at best, or even despite distrusting all candidates—but not based on trust in any of them.

Stephen Napier (Villanova University) – Can Moral Understanding be Transferred via Testimony?

Receiving knowledge from another via testimony requires epistemic trust. There are five positions regarding moral knowledge (MK) via testimony. (1) Testimony *cannot* be a source of (MK). (2) It can be a source of (MK) (Hopkins, 2007). (3) Testimony can transfer (MK) but not moral understanding where understanding involves being able to explain why-p (Hills, 2009). (4) Testimony can be a source of moral understanding. And finally, some cases are such that (5) testimony *must* be the source for moral understanding. Hills argues, persuasively, that moral understanding cannot be transferred via testimony. Consider Ron who comes to believe that it is wrong to kill an innocent person because he trusts what his Rabbi tells him. Ron has a true belief, but he fails to resonate with the core truth-maker for that belief, namely the innocent person's *inviolable right* to life (Hills, 2009, 115 ff.). Hills is right that Ron does not understand why it is wrong to kill an innocent person; 'the Rabbi said so' is not why it is wrong to kill the person.

In this paper, I present an argument for (4) – arguably a counter-intuitive position. Briefly, I present a series of cases in which an agent comes to understand another's emotional and conative state. A beloved comes to believe that her suitor loves her based on his say-so, as depicted in Verdi's *La Traviata*. A non-fiction example includes empathy-based crisis

intervention programs for sexual assault survivors. In coming to hear what it is like to be assaulted, participants in the program come to understand how to provide crisis intervention for assault victims (Foubert, 2010). Another example includes a therapist coming to understand a patient's corrosive behavioral patterns and ways to correct them. All these cases share a feature, namely; the receiver comes to have moral understanding because a morally relevant feature of treating another lovingly can *only* be transferred via testimony. This feature is endemic to all altruistic, interpersonal relationships. The implications for an ethics of trust is that trust is parasitic on shared altruistic motives. Moral understanding of a policy can be transferred via testimony from experts if experts were transparent about their altruistic motives. One practical conclusion is that perceived or actual conflicts of interest disrupt testimonial transfer of beneficial information.

Annalise Norling & Zara Anwarzai (Indiana University) – Taking the 'Expert' out of Expert Systems

Expertise is a social process which necessarily involves joint action. Once we assign expert status to a skilled agent and they accept that status, obligations are generated. The obligations for experts in particular roles—e.g., for a physician diagnosing and treating a patient—require different kinds of expertise: A physician must not only keep his medical knowledge current and consistently give accurate, relevant information to his patients, he must also understand how medically relevant information may relate to the goals, desires, and intentions of his patients.

The recent development of chatbot diagnostic expert systems (CDES) aims to improve healthcare. CDES are becoming increasingly embedded in our communications as well as the decisions we make with one another. Progress in the development and use of CDES has led to claims that their accuracy and predictive value are the same as or greater than that of human healthcare professionals. In some instances, it seems that we are directly engaging with these so-called intelligent tools, which might further suggest that CDES could fully play the role of human agents in these exchanges.

We argue, however, that CDES are not genuine experts and, as a result, they are not capable of discharging the obligations associated with medical expertise that we typically assign to human agents. Medical expertise requires joint action, and CDES are not agents with whom joint action is possible. With diagnosis and treatment, a physician's ability to identify and communicate relevant information depends on her ability to recognize the intentions of her patients. Similarly, the success of a physician's actions depends on whether a patient receives the actions of the physician as intended. If CDES are incapable of these processes, then they are not experts, and we should rethink their role in healthcare and whether they are appropriate objects of our trust.

Gloria Origgi (Institut Jean Nicod and PERITIA) – The Duty to Trust and the Duty to be Trustful

Trust is a complex attitude that has emotional, cognitive, and moral dimensions. A difficulty to reduce trust to a simple emotional attitude is that trust raises normative pressures: if someone asks you to be trusted you feel the normative pressure of not letting him or her down, and if someone trusts you, you feel the normative pressure of honoring his or her trust. These normative pressures seem to have an irreducibly social character: pressures are effective insofar as they may raise emotions of shame in those who violate the norm of trust and resentment and contempt in those who are victims of the violation. In this paper, I will investigate the relation between the affective dimension of these normative pressures and their moral dimension by arguing that an important moral asymmetry exists between the duty to trust and the duty to be trustful.

Silvia Caprioglio Panizza (University of Pardubice) – Trust and the paths we don't go down

In a dual effort to re-define trust and to broaden its application to inanimate objects, C. Thi Nguyen (2020) has offered a definition of trust as an 'unquestioning attitude'. In response to worries about the applicability of trust, rather than reliance, to objects (e.g. Baier 1986), Nguyen invokes the fact that such absence of doubt is found in trust and not in reliance, and explains the normative dimension by construing trust in objects as in something *integrated* into our own thinking and functioning. This, for Nguyen, satisfies Holton's (1994) demand that trust, as opposed to reliance, warrants betrayal. In this paper I present some worries about such articulation of trust in objects as a form of self-trust, while I build on Nguyen's suggestion of placing absence of questioning at the centre of trust by drawing on the ethical domain of moral impossibility. In such cases, what is unconsidered (here, the possibility of questioning and doubt) remains outside the scope of an agent's range of possibility for moral reasons – overt or covert. By spelling out cases in which trust and moral impossibility overlap, I aim to offer further reasons for the normative dimension of trust and bolster its distinction from reliance. In particular, I will focus on cases in which questioning scientific information from trustworthy sources amounts to going down paths that are morally and epistemically dangerous at the same time.

W. Jared Parmer (RWTH Aachen University) – Manipulation as Covert Non-Cooperation

Cooperation is a mainstay of life, especially in multidisciplinary collaboration between experts. One important function of such cooperation is to enable us to work together to be responsive to reasons, where as individuals we are in no position to understand each of the relevant normative reasons. For reasons-responsiveness to be distributed across groups in this way, fully cooperative participants must exhibit an ongoing openness to one another, such that the joint specification and reconsideration of plans is meaningfully shaped by each other's practical point of view. This renders fully cooperative participation aspirational in the following sense: each participant's concern (for fully cooperative activity) cannot be given much content in advance. Manipulators exploit this aspiration by offering and scrutinizing proposals for decisions, which normally expresses the (mutual) concern for fully cooperative activity. In this way, however, manipulators thereby mislead others about their lack of such concern. Thus, on the account I develop here, manipulation is covert non-cooperation.

Multidisciplinary collaboration among experts also shows why certain extant norm-based theories of manipulation fail (such as those offered by Moti Gorin and Michael Klenk). These theories say that the following norm is constitutively violated in manipulation: to see to it that, in influencing another person, one is responsive to their normative reasons. Now, this norm is rather demanding: it requires that, for example, a prop manager for a film-shoot be motivated by the prop-making metalsmith's normative reasons, which in turn requires that the prop manager appreciate the normative force of the metalsmith's reasons. But now we can see that it is overly demanding: experts collaborate across disciplinary boundaries precisely because each expert is not in a position to appreciate every other expert's normative reasons. And, of course, none of this need be manipulation, so violating this norm is not constitutive of manipulation.

Yevgenya Jenny Paturyan & Sara Melkonyan (American University of Armenia) – Revolution, Covid and War in Armenia: Impact on Various Forms of Trust

In the past four years Armenia experienced three major events: the Velvet Revolution (2018), the Covid-19 pandemic and a devastating war with Azerbaijan (2020). The first one boosted trust towards government, while the Covid and particularly the war undermined trust and created a sense of deep disillusionment. Nonetheless, the Armenian government weathered the post-war political turmoil and renewed its democratic mandate in snap parliamentary elections in June 2021, making it an exceptional case of a government that lost the war but got re-elected.

This paper seeks to understand how various crises (the popular uprising of 2018, the pandemic and the war) impact various types of trust in Armenia. More specifically, we look at

interpersonal trust, trust towards various institutions and branches of government, trust towards experts, and general trust towards democracy.

Trust is one of the resources that allows governments to overcome crises. Which types of trust in Armenia were most impacted by both positive (democratic peaceful revolution) and negative (pandemic, war) monumental events in the country?

Using World Value Survey, Caucasus Barometer and other available surveys, the research explores trends in various types of trust in the Armenian society in 2017-2021. We try to understand how trust is impacted by events in the country, and how it is determined by individual traits, such as education, socio-economic status, political orientation, values and so on.

Gereon Rahnfeld (Bauhaus-Universität Weimar) – Which expert should we trust? On the selection processes of experts and their stabilization

As far as governance structures are concerned, the last decade has presented an ambivalent picture. On the one hand, participatory processes and democratic structures have been strengthened, for example through the introduction of citizen assemblies as in the case of the "Conference on the Future of Europe." On the other hand, experts and epistocratic structures have become more important, as demonstrated by the influence of scientists on policy-making during the Covid-19 pandemic. This raises the question not only of the democratic deficit of epistocracies but also of the epistocratic deficit of democracies – a dispute which dates back to the Dewey-Lippman debate and is still relevant today (for recent contributions, see "Jason Brennen: Against Democracy, 2016" and "Hélèn Landemore: Open Democracy, 2020").

While the two positions of the debate point to their extremes, the middle ground is often ignored: Experts in democratic and participatory processes (see for reference to this problem Cathrine Holst and Anders Molander: Epistemic Democracy and the Role of Experts, 2019). But it is in the latter context that the question ‘Which expert should we trust?’ becomes increasingly important. In participatory processes it is not only a small circle of decision-makers who need to trust the selected experts, but rather a larger group of citizens who ought to trust experts. In my paper, I will present a sociological case study of a citizen assembly that took place in Germany in 2021 on the topic of “Germany’s role in the world”. Using this example, I will analyze how experts were selected in this participatory process and how their roles had to be stabilized.

Pablo Rivas-Robledo (University of Genoa / LMU München) – Trust in Public Expertise for Public Policy

One of the central ethical problems that epistemic democracy faces today is to decide how to treat the knowledge that ordinary citizens can provide in public deliberation. This is because, on the one hand, epistemic democracy wants to determine the specific conditions under which deliberation guarantees the best solution, yet, on the other hand, it has to deal with the aggregation of judgements of citizens with diverse social and educational background. In this talk I will examine how this problem is manifested in the creation and implementation of public policy and how can be solved. I will approach the problem under the framework offered by the CrowdLaw movement.

CrowdLaw is a theory of legislation that takes public participation to the next level. It proposes that public participation should be the backbone of public policy at all its stages, which are problem identification, solution drafting, implementation and evaluation. This is done by enriching the process with meaningful and helpful participation from citizens. In turn, this will help improve the quality of public policy, given that we are drawing from the knowledge from the collective and not simply from the opinions of citizens.

Thus, CrowdLaw presents itself as a new and promising way of infusing public policy with democracy that seeks to resolve various issues of modern democracies. To do so, CrowdLaw claims that public expertise should be ubiquitous at every step of the process so that there is room for legislative excellence in democracies, because it not only offers more legitimacy to its outputs, but also the quality of the legislation is substantially improved by incorporating the testimony of public experts in the process.

Thus, in this case the problem seems to be how to treat the testimony of presumably highly competent agents. I think the treatment that CrowdLaw gives to this knowledge is problematic. Defenders of CrowdLaw talk again and again of citizen's input as *public expertise* and how ordinary citizens are experts in their own right. Then, when ordinary citizens participate in policymaking it seems that they turn from ordinary laymen to experts. In this talk, I will explain why this is highly problematic based on the (social) epistemology of expertise (impossibility of establishing expertise, biases) and the impossibility of applying positive results of social choice theory that had established superiority of the laymen over experts.

Shane Ryan (Singapore Management University) – Trust and Epistemic Coverage

Expanding on previous work, this article argues that trust relations, or the lack thereof, have important implications for belief formation beyond testimonial belief. Focusing specifically on

trust in institutional testifiers, such as scientists, academics, politicians, and journalists, the case is made that an absence of trust in those testifiers as good testifiers in their domains, leads to those who lack trust to be worse off epistemically than would be the case if they trusted and those they trusted were trustworthy. Those who lack trust in such cases are not just made worse off with respect to missing out on testimonial knowledge but also with regard to knowledge generally, as they lack the epistemic coverage that they might otherwise gain from trusting trustworthy sources. In order to make this case, this article presents a goodwill account of trusting. Next Hardwig's epistemic dependence thesis is presented to support the case that not trusting in certain testifiers in certain domains is incompatible with gaining knowledge in those domains. Finally, Goldberg's concept of coverage is used to explain how an absence of trust can lead to missing out on knowledge beyond testimonial knowledge.

Hayarpi Sahakyan (Yerevan State Medical University) – Think well, before you count on!

Trust is essential both for human society and for interpersonal relations. Without trust, human coexistence as we know it today would be impossible. Trust is a byproduct of evolution which enables us to divide epistemic and non-epistemic labor and thus to benefit from narrow specializations. Trust is also essential for formation and development of human personality.

But what exactly is trust? Trust is a relation between the two agents. One of the agents trusts: *truster*; the other agent responds to trust: *recipient of trust*.

One widespread definition of trust views it as a belief in a high probability that a person trusted will behave in ways the truster expects. In a recent paper Karen Jones shows that this definition does not capture the heart of trust. Developing her counting on theory of trust Jones conceptualizes it as a non-moral relation between people which enables individuals to directly involve other agents in one's own projects. She acknowledges that trusting people may involve estimations of probabilities of how trustees will behave, but people can trust others without any estimation of probability.

Jones views trust relations as relations between two equally active and important agents and based on this she identifies a number of principles of trust and trustworthiness that these agents must follow. In this paper, I analyze her view and show that she is right when she claims that norms grounding trust are prudential rather than moral. However she is most probably mistaken when she holds that both the truster and the trusted person are equally responsible for good trust. My analysis of trust in this paper leads me to the conclusion that norms of trust are binding only for the agent that trusts. Before counting on others, one must think well!

Regina Schidel (Goethe-University Frankfurt) – Epistemic Justice as Precondition for Trust in Expertise

Debates about the ethical implications of trust in scientific expertise often have their focus on properties that determine the trustworthiness of experts, or ask about how to trust responsibly. This approach has a strong individualistic bias, examining virtues of trust or trustworthiness both on the side of the trustor and on the side of the trusting.

In my paper, I argue that this perspective is incomplete and that we need to focus much more on the social and political preconditions for successful trust in experts. Important impulses in this regard can be taken from the debate on *epistemic injustice*, which was initiated by Fricker and other feminist and critical theorists.

The debate on epistemic injustice addresses the question of how credibility and epistemic authority are unequally distributed along social and group hierarchies. While this is an important point, Fricker and others have the tendency to frame this primarily as a problem of (in)justice. I argue that a fundamental recognition of all members of society as epistemic authorities and a certain degree of participatory trust (Medina) is not only a (somehow independent) requirement of justice, but also systematically linked to the question of trust in experts and science. Thus, the possibility of epistemic trust depends on social and political presuppositions that need to be interrogated more closely for their normativity.

I will follow the connection between epistemic justice among members of a society and trust in experts in three steps.

On a *diagnostic* level, I show how and to what extent the reflections of the epistemic justice debate on hermeneutic, testimonial, and participatory (in)justice can be made fruitful for the question of trust. This enables a *critical* perspective on trust in experts that distinguishes successful from failing (because reactionary or authoritarian) forms of trust. Finally, this opens up a *normative* perspective in which the justice-relevant preconditions for successful epistemic trust are formulated.

Mattias Skipper (National University of Singapore) – Wise Groups and Humble Persons: The Best of Both Worlds

In this talk, I address a problem that arises when we try to harness the “wisdom of the crowd” from groups comprised of individuals who exhibit a certain kind of epistemic humility. I begin by posing the problem and then make some initial steps toward solving it. The solution I arrive at is tentative and may not apply in all circumstances, but it promises to alleviate what seems to me to be a problem of both theoretical interest and practical urgency.

Carien Smith (University of Sheffield) – Conspiracy theories and the harms to scientists: a case for a pseudoscientific epistemic injustice

Beliefs in unwarranted conspiracy theories about scientific issues that could have serious harmful consequences, such as conspiracy theory beliefs about climate change and Covid-19 vaccines, have a few central features that they share. Firstly, they are about scientific issues. Secondly, these particular beliefs are in unwarranted theories, meaning that there are some epistemic failings in the formation of these beliefs. Thirdly, they are beliefs that could have serious harmful consequences. These are only a few features. The feature that I am particularly interested in is that these beliefs centre around scientists and their scientific practices, scientific legitimacy, scientific reliability, and other related notions. These all relate to epistemic practices of some sort and the status of the scientist as an epistemic agent in these epistemic practices. In the case of beliefs in such conspiracy theories about scientific matters, scientists are subject to significant distrust and animosity from large sections of the public. I believe that through this distrust, the second-guessing of the expertise and validity of the research, and animosity, there is a type of epistemic injustice towards scientists. The significance of this problem is that this epistemic injustice can significantly delay scientific progress in fields of research that have huge impacts on the wellbeing of persons: both scientists, and the public. When we consider pressing issues such as climate change and the Covid-19 pandemic, these pseudoscientific practices disrupt progress towards sustainable and urgent action to address some of the most significant problems of our time.

I propose to call this type of epistemic injustice a pseudoscientific epistemic injustice. By "pseudoscience", I broadly refer to those categories of activities and beliefs that are “A pretended or spurious science; a collection of related beliefs about the world mistakenly regarded as being based on scientific method or as having the status that scientific truths now have” (Oxford Dictionary). This paper aims not to give a comprehensive analysis or argument for what pseudoscience is, which is in itself a complex notion, but to relate those activities and beliefs typically and broadly associated with pseudoscience to a particular type of epistemic injustice.

Anders Søgaard (University of Copenhagen) – Can Machines Be Trusted?

The Artificial Intelligence (AI) literature increasingly speaks of the need for trustworthy AI. Typically this means something like AI technologies that are interpretable, exhibit limited social bias, and cannot be used for unethical purposes. What exactly, though, does trustworthy AI really mean, and what does it have to do with trust? Does the concept of trustworthy AI reduce to reliable AI, for example? Clearly not, since reliable AI does not have to be interpretable or fair. On the other hand, Karen Jones and others argue that trust is a uniquely

human relationship. Is the objective of trustworthy AI then more than reliance and less than trust, somehow? This paper sets out to argue that in fact the objective of trustworthy AI is real trust in the general sense of Jones, and very similar to the kind of trust we put or do not put in institutions. We present a more inclusive (and, we argue, consistent) reformulation of Jones' definition and argue that this reformulation is both more useful than reducing trustworthiness to reliability than Annette Baier's definition of trustworthiness in terms of 'goodwill'.

Laura Specker (Fordham University) – Climates of Trust

Trust is a well-developed ethical concept explained by competing philosophical accounts. Yet the relationship between individual instances of trust and general atmospheres of trust is undertheorized. Annette Baier, who introduces the idea of "climates of trust" but does not offer a full account, writes that: "I have alluded to… society-wide phenomena as climates of trust affecting the possibilities for individual trust relationships" (Baier 1986: 258). Baier recognizes that individual decisions to trust are not isolated events. As she notes: "If the network of relationships is systematically unjust or systematically coercive, then it may be that one's status within that network will make it unwise of one to entrust anything to those persons whose interests, given their status, are systematically opposed to one's own" (Baier 1986: 259). In short, the wisdom of individual instances of trust depends on systemic features of ones' social context, i.e., climates. Individual trust decisions cannot be evaluated in isolation from their climates; the climate can make the difference between a wise and an unwise choice.

In this presentation I provide an account of climates of trust. I propose that such climates emerge from the development of mutual understanding between people in social relationships. While these climates emerge from understanding, for their sustenance they depend on the degree of solidarity within a culture or community, determined by the extent to which people are perceived as willing to work together to achieve each other's well-being. I explain the influence of climates of trust or distrust on individual assessments of trustworthiness and decisions to place or to withhold trust. I suggest that the perception of social solidarity can influence individuals' willingness to take risks for and with each other. Finally, I consider how climates of distrust might be transformed into climates of trust.

Eugenia Stamboliev (University of Vienna) – Trust and Maintenance — From Design Norms to Bergsonian *Durée* of Technology

From content moderators in the Philippines who sort up to 25.000 google images a day to "Turkers" paid by Amazon's crowdsourcing platform who create reviews endlessly (Couldry & Mejias 2018), companies such as Google and Amazon maintain their products thanks to

exploitative and invisible labour of actual humans (Couldry & Mejias 2019; Riesewieck & Block 2018). What is sold as a sophisticated product, often masks historically continuous labour practices. Even if technology studies still pay more attention to technology dehumanising jobs or industries (Popper 2016; Halpern 2015; LeGradeur & Hughes 2017, Levy 2017, Ford 2014, etc), trustworthiness research and techno-ethical literature on trust do not include work practices into their framework in defining trust (Robinson 2020; Bidner & Francois 2010).

Considering that thinking trustworthiness of intelligent technology or online platforms now demands a holistic integration of design, implementation and verification, the aim of this paper is to focus on trust as a durable and ongoing process embedded process and as maintenance of technology. While Nickels et al. (2010) still focus on the trustworthy *artefact*, this paper disputes if trustworthy model (especially in learning systems) is *ever* finished or artefact bound. I turn to Bergson's concept of *durée* as a multiplicity and fluidity to elaborate on trustworthiness as a process of maintenance and conditions thereof. This means to rethink the norms of trust and to move beyond trust as a fixed norm, but instead, to see it as something fluid, manifold and maintained continuously. By focussing on invisible and exploitative aspects in running technologies or innovation, I pay attention to labour practices, regulatory freedom, and geopolitics of labour and to look at the integrated exploitative nature of maintaining and innovation historically.

Through aligning trust in/of technology to Bergsonian *durée* within and of maintaining technology like platform architectures, I will challenge normative categories and expand on the concept of trustworthiness important for regulatory policy. By including a socioeconomic labour critique into philosophy of technology, I raise the question on how far ethics reaches into the implementation, maintenance, and verification of technology.

Rowland Stout (University College Dublin and PERITIA) – The Ethical Responsibility to Trust

While ethical responsibilities to be trustworthy and to engender conditions of trust are quite well theorized, the question under what circumstances one has an ethical responsibility to trust, even in the absence of evidence of trustworthiness, has been less widely discussed.

Ninni Suni (University of Helsinki) – What Is the Right Response to Misplaced Distrust?

Distrust in authorities plays a key role in problematic behavior such as vaccine hesitancy (Kärki 2021). Often this distrust has roots in perceived injustices in healthcare, such as racism,

sexism, or having been silenced or questioned (e.g., Navin 2013). Disproportionate distrust in authorities tends to be met with blame and ridicule, which points to an interesting distinction: blame is a characteristic response to a perceived moral wrongdoing, but ridicule hints that others perceive the mistake as epistemic. But both blame and ridicule potentially contribute to further polarization. What is the right response to misplaced distrust? Is the mistake in it epistemic or moral?

The mistake in disproportionate distrust can be conceptualized as one of generalizing from a non-representative set – say, treating one expert's mistake as representative of the group as a whole. As such, it seems to be an epistemic mistake, which merits withdrawal of epistemic trust (cf. Kauppinen 2018). But distrust is also recalcitrant to counterevidence in a way characteristic of motivated reasoning. When so, we should seek to find out which motivations are at work. They could be morally dubious ones, such as complancency, but they could also be feelings of having been betrayed, that is, reactive attitudes to a felt moral wrong. This distinction points to two different types of motivated distrust: the first directed at someone without prior evidence of their conduct, the other based on loss of trust. While the first can be morally wrong, the second is more complex, as it has roots in a former moral wrong. Thus, proper reactions to misplaced distrust can also be divided accordingly: moral reactions to morally bad attitudes on one side, and reactions to feelings of vulnerability and betrayal on the other—attempts to re-build trust.

Yafeng Wang (Chinese Academy of Sciences) – Crisis Investigations, Trust, and Responsibility Attribution

When a crisis—be it an engineering disaster, a financial meltdown, a terrorist attack, or a pandemic—has occurred, there are often calls for an official investigation into the exact details of what had happened during the crisis, what went wrong, and how to fix the identified problems. A crisis investigation seeks to learn from the crisis and to provide closure to the victims, their communities, and the public. To serve these purposes, however, the investigation and its results must be trustworthy.

In this paper, I argue for a requirement for the trustworthiness of a crisis investigation. To be trustworthy, a crisis investigation should avoid assigning blame and moral responsibility, especially when the political or institutional stakes around the crisis are high. I make two arguments to support this claim.

First, constructing an authoritative account of the facts and causes requires extensive information sharing among stakeholders of the investigation. The prospect of being blamed and held morally responsible for the crisis, however, inhibits information sharing among stakeholders of the investigation.

Second, the epistemic authority of an account of the facts and causes of a crisis requires consensus emerged from critical discussions among a sufficiently diverse group of experts participating in the investigation. Assigning blame and moral responsibility during a crisis investigation creates an adversarial environment which reduces the likelihood of consensus among such a group of experts, especially when political and institutional stakes are high.

I further support my arguments with case studies of crisis investigations that have the reputation of being trustworthy, including a few major engineering failure investigations conducted by the National Transportation Safety Board (NTSB) in the United States.

Jörn Wiengarn (HLRS, University of Stuttgart) – Responsibly Distrusting Scientific Experts

In the recent debate on the relationship between science and the public, an increased mistrust in science is usually highlighted as a problem. While this worry is all too understandable, however, it should not be ignored that “blind“ or uncritical trust in scientific experts is not a desirable attitude either. The ideal relationship of the public to science therefore seems to lie somewhere in between these two extremes of blind trust and blanket mistrust.

To better grasp this “middle between two extremes“, it stands to reason to investigate the specific conditions under which trust or mistrust in science is more or less appropriate. However, in order to be able to tackle this task, it seems to me that a conceptual problem arises first, which will be the focus of my talk. The problem is that it is not clear what precisely is meant by “mistrust“ at all.

In general, one can say that the concept of mistrust has been strongly neglected in the literature. This has, so the guiding idea of my talk, left the concept ambiguous and obscured relevant differentiations between different forms of mistrust. I therefore propose some lines of distinction to introduce different kinds of mistrust that I consider relevant. To just give a few hints on this: It seems, for example, to make a difference whether we speak of mistrust in experts as a *suspension* of judgment whether the expert is trustworthy or a settled *belief* that she is *not* trustworthy. Furthermore, one can understand mistrust in experts as being directed at individual statements by experts, or more fundamentally at the very criteria or methods by which they arrive at such judgements in the first place.

The various forms of mistrust, I will argue, differ not only in their epistemic structure and the way they can be remedied, but also in their epistemic and practical implications and therefore in how we should evaluate them from an ethical perspective.

Tomasz Żuradzki (Jagiellonian University) – Trust in science, democratically endorsed values, and risk preferences

The simple view on trust in science is that science is trustworthy because it deals only with facts, not values. Some more nuanced views assume that trust in science is also based on some additional non-epistemic factors, like scientists' 'moral integrity' or the usefulness of their work for the benefit of society (Hendriks, Kienhues et al. 2016). Recent philosophical discussions on non-epistemic values in science highlight the whole spectrum of roles that values play in scientific inquiry, in particular in inspiring scientific questions, affecting scientific methodologies (incl. conducting statistical analyses), classificatory practices (e.g. in biomedicine or ecology), and setting the level of evidence needed for drawing conclusions or recommendations (Elliott and Richards 2017). This last role refers to the argument from inductive risk (Douglas 2009), which states that evidence gathered to test a scientific hypothesis often underdetermines whether scientists should accept or reject the hypothesis. Scientists' decisions are based on risk preferences, e.g., on how do they evaluate the trade-offs between two types of risk: accepting false positives vs. accepting false negatives. The observation that many high-profile scientific conclusions are based on value judgments may significantly undermine the claims of science to public trust.

In my presentation, I examine some recent proposals for handling this problem. Some have called for greater transparency about how values affect scientific results (Kitcher 2003, McKaughan and Elliott 2018). Other argued that adherence to the same methodological conventions is crucial for trust in science (Wilholt 2013) because scientific results can be interpretable (and thus trustworthy) to the public only if they are based on high fixed standards (John 2014). Still, others argued that distrust is (at least partially) caused by the divergence between the values accepted in scientific practice and those by the members of the public (Irzik and Kurtulmus 2019). In particular, I critically analyze the recent proposal by (Schroeder 2021), who postulates grounding scientific processes in democratic values, i.e. the values held by the public and/or its representatives that can be revealed either by some procedures (e.g., a deliberative democracy exercise), or just in opinion surveys. I criticize this approach as highly idealized and democratic only in a declarative sense (e.g., because it makes scientists responsible for 'filtering' legitimate values). Moreover, this view does not recognize different individual 'legitimate' preferences, e.g., risk preferences. I will visualize this last criticism by referring to this type of healthcare decisions (e.g., during Covid-19 and the problem of vaccine hesitancy, see Goldenberg 2021), in which there may exist inherent tensions between public health and individual interests.