Programme Booklet

of tri-annual conference

Philosophy of Science in a Forest

Thursday 31 – Friday 31 May 2025

at

Internationale School for Wijsbegeerte

Leusden

The Netherlands

organized by a committee installed by

NVWF

NEDERLANDSE VERENIGING

VOOR WETENSCHAPSFILOSOFIE

Introduction

The *Dutch Society for the Philosophy of Science* has been organizing the current conference since 1998 (1st installment); the current one is the 10th installment. Yours truly baptized it **Philosophy of Science in a Forest** in 2010 when co-organizing it (5th installment), and has been co-organizing it since then, except for the first post-pandemic one in 2022. Once the Dutch Society was big, but fewer and fewer members (paying members) led to the following change. Only a tiny group of people now constitutes the members and board to keep the Society alive; it goes into hibernation for nearly three years, and then wakes up with a boom, in order to organise **Philosophy of Science in a Forest**, or to out-source it to a committee … Due to lack of paying members, there sadly is no money to organize other activities anymore.

This year's installment is like the previous one in 2022: limited to roughly 1.5 days; it used to be 2.5 days, but current prices have made a 2.5 day version beyond any reasonable budget. We hereby thank The Beth Foundation for contributing 1 k€, which we spend on paying for the two invited speakers, and used for a slight lowering of the price for participating masters students and PhDs. We also thank the OZSW (Dutch Research School for Philosophy) to take care of the registration.

Only a week before the deadline, prospects for the conference looked dim, but then, almost miraculously, just before the deadline, we received more than enough submissions to pull it through. (Joy.) A striking fact is the large number of contributions of philosophy of social science; philosophy of natural science has nearly disappeared. (Sadness.)

We sort-of grouped similar contributions together in parallel sessions, taking wishes of various contributing speakers into account.

Converse and network away for nearly two days, fellow philosophers of science, in a forest!

F.A. Muller
13 May 2025

Conference Information

Organisation and Programme Committee

F.A. Muller, F. Russo, J.-W. Romeijn, V. Gijsbers

Website Dutch Society for the Philosophy of Science (DSPS): www.nvwf.nl .
In Dutch: Nederlandse Vereniging voor WetenschapsFilosofie (NVWF).

E-contact:
💻 philsciforest2025@gmail.com

Venue

Internationle School voor Wijsbegeerte (ISVW), Leusden, Amersfoort.
Dodeweg 8, 3832 Leusden, The Netherlands.
Website: https://isvw.nl
💻 info@isvw.nl
☎ Tel. 033 – 465 07 00 (reception)

Directions

Public transport

Bus 19 from Amersfoort Central Station (twice per hour on weekdays, not in weekends). Website of ISVW warns that during the summer, there are fewer trains going to Amersfoort due to maintenance activities of the Dutch Railways (NS).
For moral certainty, plan ahead: https://9292.nl.
NS-App: https://www.ns.nl/en/travel-information/ns-app

Car

Use: https://www.anwb.nl/verkeer/routeplanner or Google maps.

The wolf has returned in The Netherlands

Schedule

Thursday 29 May 2025

11:00 h.: Expected Time of Arrival (ETA), for checking in before the conference begins.

Time	Room 1	Room 2
11:30 h.	**Opening: F.A. Muller** (Chair Organising Committee)	
11:35 h.	Keynote: Rosa Runhardt, *Categorizations Qua Conventions: Measuring the Arbitrariness of (Social) Scientific Categorization* (Radboud University Nijmegen) **Chair: Federica Russo**	--
12:30 h.	🍽 **Lunch**	
Chair	**F.A. Muller**	**Simon Lohse**
13:30 h.	*Why Einstein May Have Had Good Reason to Oppose the Geometrization of Gravity in General Relativity,* Femke Kuiling (University of Minnesota)	*Environmental Behavior and Norms,* Jelle de Boer (Free University Amsterdam)
14:10 h.	*Presentist Velocities* Victor Gijsbers (Leiden University)	*Causal-Mechanistic Analyses in Evidence-Based Policy,* Paride Del Grosso (University of Antwerp)
14:50 h.	*A Structuralist Metaphysics for Color Experience* Lieven Decock (Free University Amsterdam)	*Universalisation and Group Agency: Distinguishing Kantian Optimalisation from Team Reasoning,* Bronagh Dunne (Free University Amsterdam)
15:30 h.	**Break**	

Chair	F.A. Muller	Federica Russo
15:45 h.	*Systematicity Naturalized: How (Not) to Solve the Demarcation Problem* Lefteris Zacharioudakis (Erasmus University Rotterdam).	*Three Problems for Predictive Policy Advice,* Simon Lohse (Radboud University Nijmegen) & Philippe van Baßhuysen (Leibniz Universität Hannover)
16:25 h.	*Science and Metaphysics: Two Sides of the Same Coin,* Reynier Pet (Leiden University)	*Objectivity of Measuring the Effectiveness of Interventions for Alcohol Use Disorder*, Saana Jukola (University of Twente)
17:10 h.	*Requirements of a Validity Concept for Cognitive Neuroscience,* Jolien C. Francken (University of Amsterdam) & Yingying Han (Radboud University Nijmegen).	*How Peer Review Panels Measure Higher Education Quality,* Gabriel Heinrichs (Groningen University).

18:00 h.: 🍽 Drinks & Dinner

20:30 h.: NVWF General Assembly

21:30 h.: 🍸 Drinks

Dutch wolf carrying a young

Friday 30 May 2025
08:00 h. 🍽 Breakfast

Time	Room 1	Room 2
9:00 h.	Keynote: Caspar Jacobs *On Mathematics as the Language of Physics* (Leiden University) **Chair: Jan-Willem Romeijn**	---
10:00 h.	Break	
Chair	**F.A. Muller**	**t.b.a.**
10:15 h.	*A New Counterfactual Theory of Actual Causation,* Lennart Ackermans (Munich Center for Mathematical Philosophy)	*Understanding and Controversy in Evolutionary Biology,* Ludo Schoenmakers (Konrad Lorenz Institute for Evolution and Cogn. Reasearch)
10:55 h.	*Complex System: a Changing Knowledge Concept since the 18th Century,* Maarten G. Kleinhans (Utrecht University)	*Clustering Models in Psychiatry: Practical Tools or Path to Natural Kinds?* Anna van Oosterzee (Utrecht University)
11:35 h.	*Nomic truthlikeness in the light of a probabilistic representation of propositions,* Theo A.F. Kuipers (Em. University of Groningen)	*Microbiomedical Research and the Myth of Ethno-racial Categories,* Aline Potiron (Freudenthal Institute, Utrecht University).
12:15 h.	🍽 Lunch	


Dutch wolves

Chair	Victor Gijsbers	Jan-Willem Romeijn
13:15 h.	*The Loss of Ideology in the Philosophy of Science,* Thijs Ringelberg (Groningen University)	*The Scope and Resolution of Neyman's Paradox,* Samuel C. Fletcher (University of Oxford)
13:55 h.	*Revisiting the Analogy between Grounding and Causation,* Martin Voggenauer (Radboud University Nijmegen)	*Decision Models and Performativity Problems,* Oyku Ulusoy (Bristol University)
14:35 h.	*What can we learn from replication efforts about scientific pluralism?* Stephanie Meirmans (University of Amsterdam)	*Measuring Psychological Attributes in the Network Framework,* Riet van Bork (University of Amsterdam)
15:15 h.	Break	
Chair	**Victor Gijsbers**	**F.A. Muller**
15:30 h.	*Broke But Not Worth Fixing? On Adjusting Effect Sizes for Research Biases,* Ina Jäntgen (Munich Center for Mathematical Philosophy).	*Medical research without Big Pharma: It's preferable, it's profitable, and it's practicable*, Hans Radder (Em. Free University Amsterdam)
16:10 h.	*HARKing and Meta-Analysis: A Novel Perspective,* Boris Kuiper (Groningen University).	*Stricter Standards for Causal Language in Social Science,* Nikki Weststeijn (University of Barcelona).
16:50 h.	*Science, Values, and the Interspecies Standpoint,* Claudia Cristalli (Tilburg University)	*Mitigating the Theory Crisis in Psychology: Why Formalization needs Theories of Explanation,* Luiza Yuan (University of Amsterdam).

Dutch sheep, after an encounter with Dutch wolves.

Abstracts

The order of the abstracts is the chronological order of the talks, overridden only by being in the same parallel session. First **Room 1**, then **Room 2**.

First Plenary Talk, Thursday 11:30–12:30 h.

Categorizations Qua Conventions: Measuring the Arbitrariness of (Social) Scientific Categorization
Rosa Runhardt

Researchers organize the world around them into categories. Their choices in *categorizing* a phenomenon of interest are never only determined by the area under consideration: there is flexibility in (amongst other things) scientists' choices of category boundaries, number of categories, and rules for determining category membership. I present an approach that treats categorization choices as conventions, and accordingly, I develop an information theoretic measure from the philosophy of social conventions to track how much flexibility a researcher has in some given area of categorization. Following work by Brian Skyrms and Cailin O'Connor, my approach uses Shannon entropy to define the level of arbitrariness of categorization.

I then argue that two interpretations of arbitrariness are salient for philosophers: a descriptive interpretation, which delivers the amount of consensus in any given area as well as changes in this amount over time; and a normative interpretation, which systematizes philosophical questions on pluralism, i.e., the intuitive notion that some areas of inquiry benefit from there being a variety of ways of categorizing. By discussing categorization as a form of clustering, I show that the normative interpretation is related to, but not reducible to, the question of whether the phenomenon under study is a kind. Contrasting the two interpretations, I will show the importance of distinguishing the degree of consensus on the one hand from the normative degree of arbitrariness on the other. This distinction further motivates quantifying conventionality, as it calls into question the idea that consensus in categorization is always a sign of scientific progress.

First Parallel Session, Thursday 13:30 – 15:30 h.
Room 1

Why Einstein May Have Had Good Reason to Oppose the Geometrization of Gravity in General Relativity
Femke Kuiling

Einstein's Theory of General Relativity (GR) has been understood by many, notably by H. Reichenbach (Reichenbach 1958), as 'geometrizing gravity.' In fact, nowadays this is the way GR is taught in university courses. But Einstein himself was a fierce opponent of this understanding of GR, as D. Lehmkuhl (Lehmkuhl 2014) has brought to the fore in a paper titled 'Why Einstein did not believe that general relativity geometrizes gravity.' He argues that Einstein instead understood GR as bringing about the unification of inertia and gravity.

Lehmkuhl distinguished two ways in which 'geometrization of gravity' can be understood: the first in the trivial sense that geometrical concepts are used in GR, and the second in the sense that gravity is ontologically reduced to geometry.

Einstein was against the claim that GR geometrized gravity as understood in the first sense because he thought that it meant just to say nothing at all because geometry is used in other areas of physics as well, and we don't say that those are geometrized.

However, Lehmkuhl argues that Einstein could have thought that gravity was ontologically reduced to inertia and that therefore he could have agreed that GR geometrized gravity as understood in the second sense. But, Lehmkuhl argues, Einstein's first aim was to show the unification of gravity and inertia, and when that was done, he no longer had the appetite for the ontological reduction, and that's why he didn't.

In a paper published in the same year, C. Lehner (Lehner 2014) has drawn attention to Einstein's 'methodological realism' and argued that Einstein was very attached to drawing a distinction between the elements of GR that do, and the elements that do not, depend on coordinates.

I argue that Lehner's discussion of Einstein's methodological realism provides a much stronger argument for why Einstein opposed the geometrization of gravity.

Einstein's 'methodological realism' led him to divide the elements of the theory into two categories: coordinate-independent elements and coordinate-dependent elements. If we apply this division to GR, we find out that spacetime geometry is coordinate independent, while gravity is not. Thus, we have found a better reason why Einstein might have argued against geometrizing gravity in the second sense: geometry and gravity are elements with a different ontological status.

Presentist Velocities
Victor Gijsbers

Should *presentists* – as presentists – be interested in instantaneous velocities? Such velocities have been the subject of philosophical debate since Zeno came up with his paradox of the arrow. More recent interest in them was sparked by F. Arntzenius (2000). What makes instantaneous velocities philosophically mysterious is that on the one hand, they are supposed to be qualities had by an object at a single instant of time; but on the other hand, they appear to have a necessary – perhaps even a logically necessary – connection to the position of that object at other times. This duality of being confined to an instant and being spread out over time requires explanation.

Presentists, who are often found struggling with the problem of past truths, have good reason to take a hard look at instantaneous velocities. What the presentist needs is a full description of the universe that is, on the one hand, limited to an instant of time – the present; but that is, on the other hand, rich enough to ground truths about the past, and possibly the future as well. This is precisely the form that instantaneous velocities take. So perhaps there are ways of thinking about velocity that can benefit the presentist.

In this talk, I first survey the main positions in the literature. All of these react to the observation that when we admit both velocities and positions at all times into our fundamental ontology, we admit too much: since velocities are time derivatives of positions, velocity facts can be derived from position facts. The orthodox at-at theory of velocity holds that velocities are not ontologically fundamental. An alternative theory proposed by C.D. McCoy (2018) holds that velocities are fundamental, while position facts are derivative. Both theories imply an eternalist metaphysics of time: the fundamental ontology of the world encompasses either positions at all times or velocities at all times. I propose an account of velocities that fits better with presentism, namely that what is fundamental are the positions and velocities of objects in the present moment only. Past and future velocity facts are ontologically derivate. I show that this theory is more ontologically parsimonious than its rivals, which gives us a good reason to accept it; and I defend the new and radical, but also pleasingly coherent, form of presentism that it entails.

References

Arntzenius, Frank. 2000. 'Are There Really Instantaneous Velocities?', Edited by Sherwood J.B. Sugden. *Monist* 83 (2): 187–208.

McCoy, C.D. 2018. 'On Classical Motion', *Philosopher's Imprint* 18.

A Structuralist Metaphysics for Colour Experience

Lieven Decock

Structuralism is a widely accepted metaphysical position, in particular in philosophy of language, philosophy of physics, and philosophy of mathematics. The common elements in structuralist metaphysical theories are anti-atomism — the idea that relations are more fundamental than entities — and the belief that individual entities only exist as nodes within a global structure of relations. It is remarkable that only recently structuralist views have been explored in the philosophy of the cognitive sciences and the philosophy of mind.

I will put forward a structuralist color ontology. I propose a structuralist account of color experience that challenges the conventional view, which describes color phenomenology using the linguistic subject- predicate or the metaphysical object-property format. According to this structuralist view, color shades are determined by the totality of proximal luminance and spectral differences in the visual field. This totality determines further structures, such as the similarity relation between color shades at separate locations in the visual field and the structure of color space.

To buttress this view, I draw on physiological and phenomenological data. In contemporary philosophy of color, color constancy often takes center stage, while color contrast phenomena are often neglected. To illustrate the significant role of color contrast, I briefly discuss the make-up of the human retina, the Jameson-Hurvich opponent process theory, and phenomenal evidence, such as desaturated colors, simultaneous color contrast, the Cornsweet illusion, the Ganzfeld effect, and Churchland's chimerical colors. Drawing on Whittle's experiments, I highlight the role of holistic principles such as anchoring and scaling, that illustrate that the experienced color shade of a particular spot in the visual field is determined by the totality of spectral differences in the visual field.

Within this structuralist framework, I offer preliminary explanations of several color phenomena. Metamerism depends on the discriminatory capacity of the retina. Perceptual variation, the fact that a partially sunlit cup is veridically experienced as having two colors, is not problematic, as the physical properties of the cup can easily be inferred. Relative color constancy depends on cone contrasts in the retina. Color categorization can be explained on the basis of structural features of color space. Furthermore, I offer evidence that contemporary neuroscientific research on color experience implicitly aligns with a structuralist approach. I conclude that structuralism provides a more suitable ontological framework for describing color experience than the orthodox atomist view.

References

Decock, L. & Douven, I. (2012). Qualia compression. *Philosophy and Phenomenological Research* **87**(1):129-150.

Isaac, A. (2013). Quantifying the subjective: Psychophysics and the geometry of color. *Philosophical Psychology*, **26**(2):207-233.

Isaac, A. (2014). Structural realism for secondary qualities. *Erkenntnis* **79**:481-510.

Lyre, H. (2022). Neurophenomenal structuralism. A philosophical agenda for a structuralist neuroscience of consciousness. *Neuroscience of Consciousness* 2022(1):niac012.

Morrison, J. (2020). Perceptual variation and structuralism. *Noûs* **54**(2):290-326.

Whittle, P. (2003). Contrast colours. In: R. Mausfeld & D. Heyer (eds.), *Colour perception: Mind and the physical world* (pp. 115–138). Oxford University Press.

Wolf Midas,
Never leaves Duck Town

First Parallell Session, Thursday 13:30 – 15:30 h.
Room 2

Environmental Behavior and Norms
Jelle de Boer

Recently Triodos Bank and the Social Cultural Planning bureau have written fairly optimistic reports about the potential in the Dutch society to reach a tipping point in the direction of more sustainable behavior. Studies by behavioral scientists on pro-environmental behavior show a strong effect of perceived social norms and peer effects. For example, any extra solar panel in a neighborhood increases the probability that a next solar panel will be installed. The aim of my paper is to understand the dynamic between social norms and pro environmental behavior.

In the literature on environmental behavior, scientists often distinguish between:

- descriptive norms: how people are currently behaving, and
- injunctive (prescriptive) norms: how people commonly find that one should behave.

Both of these types of norms exert a causal influence on people's behavior.

The behavioral scientists also distinguish types of values or goals that people have: hedonic goals, gain goals (money, esteem), *biospheric* goals.

In this paper, I want to investigate how these various factors hang together by integrating them into one framework and model.

The prescriptive solution is a possible action profile in a game, but not necessarily an equilibrium. And if it could be a pure equilibrium, most of the time (in our current world) it will have a very narrow or even nonexistent basin of attraction.

But a population can gradually move towards the ideal (or prescriptive norm) of full cooperation, if it starts out above a certain favorable threshold (the level is given by the descriptive norm), and when individuals are partly motivated by 'environmental guilt', and by a belief about how many people around them are acting pro-environmentally. If many of them are not acting so, they will feel excused. The dynamic goes like this. A lower level of excuse induces a stronger sense of environmental guilt, which induces more cooperation, which induces again a lower level of excuse, which induces more cooperation, etc. More specifically, I present a model of how a moral sentiment, which I call 'environmental guilt', can develop endogenously in a group of people, depending on each other's expectations and past behavior. The approach of the paper is explanatory, not normative, and in Humean fashion.

Causal-Mechanistic Analyses in Evidence-Based Policy
Paride Del Grosso

Evidence-Based Policy (EBP) is the idea that policies should be developed on the basis of the best available evidence, since an evidential basis increases the chances that policies turn out to be effective. Effectiveness is, de facto, a causational issue. Indeed, asking whether a policy (*P*) will produce the intended outcome (*O*) can be formulated in causal terms, namely: "Will *P* cause *O*?". As such, evidence of causation – i.e. evidence showing a causal relation between *P* and *O* – plays a crucial role. To gather this type pf evidence, meta-analyses and randomised controlled trials (RCTs) are generally preferred, since these two methods are strongly reliable in showing a causal link between two variables (despite not necessarily showing how these variables are associated). Nevertheless, these methods are also often difficult to employ – because of, for instance, high costs or ethical obstacles.

In this paper, I will argue that, when meta-analyses or RCTs are not available, policymakers should combine evidence of correlation and mechanistic evidence – i.e. evidence of the existence of a causal mechanism (or more) between *P* and *O*. This is in line with the framework of Evidential Pluralism (see, e.g., Shan and Williamson 2023), the epistemological thesis according to which, in order to establish a causal relation (in this case, between *P* and *O*), two causal claims must be established first: a correlation claim (for which evidence of correlation is needed) and a mechanistic claim (for which mechanistic evidence is needed). I will support my claim by referring to a well- known example of a failed policy, namely the California Class Size Reduction (CSR). CSR was implemented in 1996 and – largely inspired by the recent positive results ofthe similar STAR Project in Tennessee – was aimed at improving students' performance in national standardised tests by reducing class size. Differently from the Tennessee STAR, CSR did not bring about the intended outcome: although students' performance improved, this was not caused by the reduction of the class size. As a consequence, the policy ended up being a significant waste of resources.

By means of counterfactual reasoning, I will explain why, if a causal-mechanistic analysis (CMA) had been employed, the failure of the policy could have been prevented.

From this example, I will generalise my claim and argue that employing CMAs is crucial for two aspects of policy development:

(i) Having a good definition of the problem that we want to solve;
(ii) Understanding whether and how *P* will produce *O*.

Concerning (i), providing a good description of the problem at hand requires, inter alia, understanding its causes and a CMA can help do so. This is crucial to then develop an effective policy. Concerning (ii), a CMA can help identify how *P* will produce *O*, as well as some potential side effects that *P* might bring about. Lastly, I will argue that CMAs should be employed in EBP, as they increase the overall reliability of causal inferences for policy decisions, improving so both policy effectiveness and efficiency.

Reference

Shan, Y., & Williamson, J. (2023). *Evidential Pluralism in the Social Sciences* (1st ed.). Routledge.

Universalisation and Group Agency: Distinguishing Kantian Optimalisation from Team Reasoning
Bronagh Dunne

Game theory models how rational agents make choices. It predicts behaviour in strategic situations and provides normative standards for rational play. A recognised limitation of standard game theory is its difficulty in accounting for cooperation and coordination in collective action problems. Classic games like the Prisoner's Dilemma and Hi-Lo illustrate this issue: a mutually advantageous strategy exists, yet standard game-theoretic reasoning deems its pursuit individually irrational. Despite this, many people contradict these predictions in experiments and real-life situations.

Addressing collective action problems within the framework of individual rationality is challenging. For this reason, a branch of the literature devoted to explaining cooperation has moved beyond the individualistic methodological framework. The team reasoning literature notes a significant difference between an individual acting alone and a group acting together as a team. To reflect this, it modifies the unit of agency of classical game theory, allowing groups to be agents in their own right with team goals and preferences (Sugden 2003, Gold and Sugden 2007). This changes the question agents ask themselves from "What should I do?" to "What should we do?". However, a recent theory by John Roemer seems to suggest that this move away from individualist methodology is too hasty (Roemer 2010, 2015, 2019). Roemer's "Kantian optimisation" takes the individual as the locus of agency and instead assumes the universalisation of action. The individual agent asks themselves, "What would I like that all did?". The agent considers what action she would take, given the assumption that all others will take the same action. This move is significant as it raises questions about the unit of analysis required to explain cooperation and the normative assumptions necessary to sustain cooperative behaviour. Moreover, the notion of group agency is controversial. Methodological individualists contest the existence of group intentions, preferences, and goals that are somehow independent of individual intentions, preferences, and goals. Kantian optimisation is novel in that it sidesteps this problem, presenting an individualistic theory of rational cooperation for non-cooperative games.

Shedding light on how these two theories might relate is relevant. And yet, it has largely been neglected, with the exception of Natalie Gold, who suggested that Kantian optimisation might be better understood as a special kind of team reasoning (Gold 2021). In other words, Kantian optimisers could be acting as team members and reasoning collectively. This implies that Kantian optimisation fits within the group agency framework and takes the team as the unit of analysis. This paper investigates whether Kantian optimisation could be understood in terms of team reasoning. I will argue against this. The two theories hinge on distinct principles. The analysis demonstrates that incorporating group agency into Kantian optimisation undermines its universalisation principle and, further, that imposing universalisation would make group agency redundant. While these two approaches may sometimes lead to the same equilibrium, one does not boil down to the other; different principles drive the underlying reasoning.

References

Bacharach, Michael. 2006. *Beyond Individual Choice: Teams and Frames in Game Theory*, ed. Gold, N. and Sugden, R. Princeton, NJ: Princeton University Press.

Gold, Natalie. 2012. *Team reasoning, framing, and cooperation. In Evolution and rationality: Decisions, co-operation and strategic behaviour*., eds. Samir Okasha, Ken Binmore Cambridge University Press
Gold, Natalie, and Robert Sugden. 2007. Collective intentions and team agency. *Journal of Philosophy* **104** (3): 109-37.
Gold, Natalie. 2021. *How We Cooperate*, John E. Roemer. Yale University Press, 2019, 248 pages." *Economics and Philosophy* **37** (2): 309–15.
Roemer, John E. 2019. *How we Cooperate; A Theory of Kantian Optimization*, Yale University Press.
Roemer, John E. 2015. Kantian Optimization: A Microfoundation for Cooperation, *Journal of Public Economics* **127** (Juli): 45-57.
Sugden, Robert. 2003. The logic of team reasoning, *Philosophical Explorations* **6.3**: 165–681.
Sugden, Robert. 2000. Team Preferences, *Economics and Philosophy* **16.2**: 175–204.

Second Parallel Session, Thursday 15:45 – 17:55 h.
Room 1

Systematicity Naturalized: How (Not) to Solve the Demarcation Problem
Lefteris Zacharioudakis

Can we account for the diversity of scientific knowledge, while at the same time preserving what separates it from other things and activities that also claim to be forms of knowledge? Paul Hoyningen-Huene (2013) answers both of these demands affirmatively. Science – understood broadly as all academic pursuit of knowledge – is characterized by *increasing systematicity* along nine dimensions. Conversely, *other putative knowledge* fails to achieve such systematicity. This account received a fair amount of attention upon publication (see Lohse, Bschir and Chang 2019) and has been consistently cited in reference articles (Hansson 2021; Hepburn and Andersen 2021; Cat 2024). Despite this uptake, Hoyningen-Huene's systematicity theory is mired by its *lack of parsimony* and its *overly descriptive* approach, impairing the presentation of its argument, and depriving it of its normative bite in actually demarcating science from non-science, respectively. Call these (a) *the challenge from parsimony*, and (b) *the epistemic demarcation challenge*.

Turning to some prominent critics, Naomi Oreskes (2017) has criticized Hoyningen-Huene for failing to account for the problem of *facsimile science*, that is, knowledge that achieves systematicity yet is not scientific. Stathis Psillos (2016, 2017) has been more severe, in that he has accused systematicity theory for failing to achieve *epistemic warrant*, i.e. an inability to show how science accomplishes *truthfulness* by relating to *evidence*. On the face of it, one could interpret both of the above critics as taking up (b), as their criticism amounts to an accusation of failing to account for the distinctiveness of science. Yet, if one reads Hoyningen-Huene's own reply to his critics (2018), one is struck by his repeatedly resorting to the comprehensiveness of his systematicity theory to defend himself. Thus, it appears that, in effect, critics inadvertently also – and perhaps primarily, if we respect Hoyningen-Huene's authorial intent – take up challenge (a), by failing to interpret systematicity theory precisely *because of* the confusing overabundance of dimensions of scientificity that it proposes. Before proposing my emendation of systematicity theory, let me turn to one more challenge, which while more subtle, might be still more devastating.

In demarcating science from *pseudoscience,* Maarten Boudry proposes that: "Territorial demarcation deals with classifications within the human web of knowledge, whereas normative demarcation distinguishes between real and false knowledge, between theories and practices that are valuable and those that are not" (2022, p. 86). Boudry cites systematicity theory as his prime example of territorial demarcation, and hence as the kind of demarcation theory that we ought to abandon if we are to address the more pressing concern of detecting and avoiding pseudoscience (ibid, passim). In effect, Boudry is proposing (c) a *sidestepping challenge* to systematicity theory.

To address challenges (a), (b) and (c), I argue that systematicity theory needs to be revised in favor of a more hierarchical scheme, while retaining its current dimensions. Combining dimensions 6 and 7 ("epistemic connectedness" and "the ideal of completeness"), I advocate for *complete epistemic connectedness* as the unifying systematic ideal, and as *a unitary demarcation criterion* in science, as against Boudry's distinction.

References

Boudry, M. 2022. Diagnosing Pseudoscience – by Getting Rid of the Demarcation Problem, *Journal for General Philosophy of Science* **53**, 83–101.
Cat, J. 2024. "The Unity of Science". *Stanford Encyclopedia of Philosophy*.
Hansson, S.O. 2021. Science and Pseudo-Science, *Stanford Encyclopedia of Philosophy*.
Hepburn, B. and H. Andersen. 2021. Scientific Method, *Stanford Encyclopedia of Philosophy*.
Hoyningen-Huene, P. 2013. *Systematicity: The Nature of Science*. Oxford: Oxford University Press.
Hoyningen-Huene, P. [2018] 2019. Replies, *Synthese* **196.3**, 907–928.

Lohse, S., Bschir, K. and H. Chang. 2019. *Synthese* **196.3**, *Special Issue on Systematicity: The Nature of Science?*.
Oreskes, N. [2017] 2019. Systematicity is necessary but not sufficient: On the problem of facsimile science, *Synthese* **196.3**, 881–905.
Psillos, S. 2016. Having science in view: General philosophy of science and its significance, In P. Humphreys (Ed.), *The Oxford Handbook of Philosophy of Science* (pp. 137–160). Oxford: Oxford University Press.
Psillos, S. 2018. Systematicity Without Epistemic Warrant?, *Journal for General Philosophy of Science* **49.1**, 127–135.

Science and Metaphysics: Two Sides of the Same Coin
Reynier Pet

One of the goals, if not the main goal of science is the accumulation of knowledge. This is hardly controversial. But it would seem that the mere accumulation of (justified) true beliefs does not adequately correspond to this goal. Scientists spend time searching for and testing general laws rather than merely recording as many empirical observations as possible. The type of knowledge science seems particularly interested in, is general, abstract. Perhaps for the same reasons, many philosophers have focused on scientific explanation rather than knowledge, or even understanding as distinct from explanation (de Regt 2017). I propose that this interest of science is not merely based on covering some kind of wider range of knowledge, but that it has value as practical knowledge. Our interest in developing theories in nuclear physics is that it allows us to do things such as build nuclear reactors. They are means to ends.

As a practical ideal, abstraction is valuable in how it provides more general means for more general goals. As such, the perfect form of such a thing would be a general law with universal applicability, a law that is always true yet could turn out to be false with every new observation. The practical value of a scientific theory exists in this space where it is as refutable as possible, while remaining as unrefuted as possible. The effort of science is therefore a balancing act between these things. Whenever we experience unexpected outcomes, we are faced with a choice to refute (at least) one of our base assumptions. Since we are loathe to give up on successful general means to ends, we, contra Popper (1959), generally give up on more specific beliefs rather than more general ones (unless an appropriate alternative is available). But this sacrifices the applicability of the theory.

However, at the same time that we engage in this effort of putting up claims and theories to be potentially refuted, we are engaging in the effort of expressing the hierarchy between those that remain. The less refuted a theory, the more fundamental it must be. Yet it must follow from this that the most fundamental of these must be the least refutable. To allow one claim to be refuted is the same act as not allowing the others to be. This leads to the perhaps somewhat surprising conclusion that engaging in science is at the very least also implicitly engaging in metaphysics.

Science and metaphysics are thus two sides of the same coin. Where the scientific side of the activity is focused on determining the truth of things, the metaphysical side is focused on how these truths should be generalized and abstracted from. We are not merely out to know things, we are out to organize what we know and might come to know in a way that provides us with the most general means to the most general ends, which is as much a metaphysical endeavor as it is a scientific one.

References

Popper, Karl. 1959. *The Logic of Scientific Discovery*. London: Hutchinson.
Regt, Henk W. de, 2017. *Understanding Scientific Understanding*. Oxford University Press.

Requirements of a Validity Concept for Cognitive Neuroscience
Jolien C. Francken

The question of 'construct validity' – whether a measurement procedure actually measures what we intend to measure – is seldom discussed in cognitive neuroscience, in contrast to the adjacent field of psychology. This is surprising, given the fact that cognitive neuroscience aims to study unobservable, mental properties, i.e., cognitive capacities such as memory or attention. Here, we explain why the negligence of concepts and measurement is problematic, and how a novel concept of validity for cognitive neuroscience could address this issue.

Neuroscientists take great care in designing controlled experiments to allow for causal/valid inferences (i.e., internal validity), and they sometimes also acknowledge the diXerences between studying a cognitive capacity in a lab setting compared to a natural environment (i.e., external or ecological validity). However, they generally do not pay attention to construct validity, which relates to inferences from measurement to higher- order concepts (1). As a result, cognitive neuroscience faces serious obstacles for data integration as well as for interpretation of neuroscientific research in a wider social and everyday context (2).

We argue that because cognitive neuroscience is characterized by several special features, a concept of construct validity cannot be simply imported from adjacent scientific fields. Therefore, we discuss requirements of a validity concept tailored to the practices of cognitive neuroscience.

First, a validity concept should be compatible with both realist and anti-realist conceptions of cognitive capacities. Neuroscientists often (implicitly) adopt a realist attitude to cognitive capacities, i.e., they believe that cognitive capacities can be found in the brain. However, it has been argued that this position is problematic, because neuroscience has shown that there is often no clear (one-to-one) mapping between cognitive capacities and neural processing (3). Recent proposals from philosophy of psychology provide an interesting perspective on this problem by introducing a distinction between the validity of the measures and the ontological status or legitimacy of the constructs that they measure (4,5). We explore the usefulness of this distinction for a concept of validity in the context of cognitive neuroscience.

Second, cognitive capacities are often ill-defined, multifaceted and context dependent. A concept of validity should therefore be compatible with such phenomena. We argue that this requires a diXerent view on validity and validation practices. In brief, validity comes in degrees (4,5) and validation is an iterative process of concept development and measurement development (6,7,8)

Third, a validity concept for cognitive neuroscience should be tailored to the specific experimental practices of the field. First, so-called experimental 'conditions' are created by the researcher to let a participant engage a cognitive capacity – e.g., working memory – more or less, allowing to study associated diXerences in neural processing. Second, cognitive neuroscience uses experimental manipulation and measurement across diXerent levels, e.g., behaviour and diXerent biological levels of the brain.

We conclude that a concept of validity in cognitive neuroscience is needed, and that existing concepts of (construct) validity are not well-suited to use in this specific context (9). Our analysis oXers useful guidance towards developing a novel concept of validity for cognitive neuroscience.

References

1. Jiménez-Buedo, M. & Russo, F. Experimental practices and objectivity in the social sciences: re-embedding construct validity in the internal–external validity distinction. *Synthese* **199**, 9549–9579 (2021).
2. Francken, J. C. & Slors, M. Neuroscience and everyday life: Facing the translation problem. *Brain Cogn* **120**, 67–74 (2018).
3. Francken, J. C. & Slors, M. From commonsense to science, and back: The use of cognitive concepts in neuroscience. *Conscious Cogn* **29**, 248–258 (2014).
4. Feest, U. Construct validity in psychological tests – the case of implicit social cognition. *Eur J Philos Sci* **10**, (2020).
5. Stone, C. A Defense and Definition of Construct Validity in Psychology. *Philos Sci* **86**, 1250–1261 (2019).
6. Feest, U. What exactly is stabilized when phenomena are stabilized? *Synthese* **182**, 57–71 (2011).
7. Sullivan, J. A. Construct Stabilization and the Unity of the Mind-Brain Sciences. *Philos Sci* **83**, (2016).
8. Francken, J. C., Slors, M. & Craver, C. F. Cognitive ontology and the search for neural mechanisms: three foundational problems. *Synthese* **200**, 1–22 (2022).
9. Han, Y. Multiple Historic Trajectories Generate Multiplicity in the Concept of Validity. *Perspectives on Science* **32**, 488–517 (2024).

Polite Dutch farmer, posting a Thank-You note on a tree.

Second Parallel Session, Thursday 15:45 – 17:55 h.
Room 2

Three Problems for Predictive Policy Advice
Simon Lohse & Philippe van Baßhuysen

Scientific policy advice offers a systematic method to enhance the rationality, trustworthiness and (societal) responsiveness of policy-making. Today, it can be seen as an integral element in progressive legislation, government work and policy-making in general and is used as a matter of course in almost all policy areas. Policy advice can take different forms, including providing input for objective setting, monitoring of policy interventions, and making predictions to guide policy decisions.

In this talk, we focus on predictive policy advice, which we discuss using examples from recent practice in pandemic, ecological and economic policy. We seek to make progress by identifying three key problems for predictive policy advice and discussing their interplay as well as possible ways of coping with them. The *first problem* revolves around the question of *epistemic pluralism*. There is broad agreement that pluralism is useful for predictive policy advice, as diverse perspectives facilitate a richer description of complex real-life problems and allow for interactive error correction (Mitchell 2020; Bschir & Lohse 2022). However, the question arises as to which disciplines should – and indeed can – be drawn upon in providing input for predictive policy advice. We explore field-specific differences in predictive capabilities, unearthing practical and normative challenges for science-informed policy.

The *second problem* concerns navigating inherent but often *implicit value influences* in predictive policy advice, which can hardly (if ever) be avoided (Carrier 2021). We address issues of value transparency and openness to highlight shortcomings in current practice, some of which, we argue, are caused by a romanticised ideal of value-free policy advice. Finally, we discuss the issue of *model performativity* as a deep problem for predictive policy advice. We focus in particular on self-defeating effects of predictive models and argue that current attempts at managing these effects fall short of providing well-developed strategies to manage the practical impacts of these effects (Khosrowi 2023; van Basshuysen 2023).

In the last part of the talk, we explore connections between the introduced issues. This not only serves to deepen the philosophical analysis, but will also open up ways to better deal with the problems. We show that pluralism, on the one hand, exacerbates the issue of value- laden predictive policy advice and, on the other hand, can provide the basis for a more rational approach toward this issue. Furthermore, we argue that value transparency and epistemic pluralism can be helpful in managing the performativity effects of prediction models. For example, reflecting on value-laden decisions regarding the endogenisation of behavioural responses in model predictions allows for a more sensible approach to addressing normative questions in this context. We close with some suggestions for improving predictive policy advice in practice. For one thing, we want to advocate promoting a less idealised public understanding of predictive policy advice that pays more attention to the discussed conceptual problems and limitations. In addition, we make the case for pro-actively institutionalising more sophisticated ways of knowledge integration in science-informed policy.

Objectivity of Measuring the Effectiveness of Interventions for Alcohol Use Disorder
Saana Jukola

Given widespread mortality and morbidity related to alcohol consumption, research-based measures to intervene on problematic alcohol use are needed. The so-called evidence- based approach, according to which decisions about the treatment of patients in clinical and social care context should be based on "conscientious, explicit, and judicious use of current best evidence" (Sackett et al. 1996, 71) has been adopted in the management of alcohol-related diseases since 1990s (Bergman & Hübner 2016). According to the tenets of evidence-based medicine and practice (EBP), randomized controlled trials and meta- analyses provide the highest quality evidence as they are the best means for securing the objectivity of the results by excluding the need for judgments from the process. Despite the criticism this position has received, EBP guidelines govern the availability of treatment options and allocation of resources in many countries (Bergman & Hübner 2016; Klingeman & Storbjörk 2016).

This talk addresses some epistemological and conceptual problems that arise when the effectiveness of interventions for problematic alcohol use, 'Alcohol Use Disorder' (AUD) in particular, are tested. I focus on the following questions: How do value-laden background assumptions involved in the conceptualization of AUD influence how the effectiveness of different interventions is established? Does the value-ladenness of trial construction challenge some of the foundational principles of EBP?

First, I argue that because AUD is multiply realized due to its polythetic diagnostic criteria, there is considerable leeway considering the composition of study population included in the trials. Moreover, researchers rely on value-laden premises concerning the disease definition and diagnostic criteria when designing trials. I illustrate the challenges that follow from this by discussing the development and testing of mHealth technology, e.g., apps, for treating and managing AUD. I show how researchers struggle with defining what condition the interventions are supposed to target and choosing what outcome measures to use.

Secondly, I argue that dealing with this value-ladenness inherent in construction of the trials excludes the possibility of achieving objectivity in the sense emphasized by the promoters of EBM, namely in the sense of excluding the need for making judgments from the process. Consequently, there is an inherent tension in testing interventions for AUD in the EBP framework.

How Peer Review Panels Measure Higher Education Quality
Gabriel Heinrichs

Higher education quality (HEQ) is generally considered too ambiguous a concept to define adequately (Weenink, 2025). Nevertheless, peer review panels have been measuring the quality of Dutch higher education programmes and institutions for decades. Moreover, since the 1990s these panels have done so while bridging different stakeholder perspectives on HEQ. How do these accreditation panels navigate these complex measurements? The panels' social deliberations play an important role, as I will argue. In my analysis of this measurement practice, I will draw on concepts from philosophy of measurement, judgment aggregation theory, and on my own qualitative studies into the Dutch accreditation system.

HEQ as a Ballung *concept.* Since HEQ is a socially constructed concept, any attempt to demarcate it into a clear and fixed set of 'essential' attributes is ill-advised. Instead, HEQ should be perceived as a *Ballung* concept which is characterised in a 'fuzzy' way (Cartwright & Runhardt, 2014). Importantly, the characterisation of such Ballung concepts should accommodate both epistemic and non- epistemic aims, as recent philosophy of measurement has increasingly shown. Psychometric measurements, for example, should strive for epistemic virtues like validity and reliability, but also align with ethical and social aims of stakeholders like patients, clinicians and insurers (Rodriguez Duque et al., 2024).

In the case of measuring HEQ, the epistemic and non-epistemic aims of academics, students, employers, and others are typically taken into account.

(Non-)epistemic iteration of HEQ measurement. Stakeholders of measurement practices should be expected to actively pursue their epistemic and non-epistemic aims. At the aggregate level, this pursuit propels a measurement practice's development over time. According to Hasok Chang, this development materialises as a succession of increasingly better –or at least: adapted– measurement instruments and underlying theories. Chang calls this development *epistemic iteration* (Chang, 2004), focussing on successive approximations of epistemic virtues within a measurement practice. But as recent philosophy of measurement suggests, both non-epistemic aims and virtues are aspired to by measurement practices' stakeholders. Therefore, measurements' epistemic and non- epistemic iteration should be considered in tandem.

Indeed, the history of the Dutch accreditation system shows that a combination of epistemic and non-epistemic aims have driven the system's iterative development since the

1980s. This can be gleaned especially from the system's consecutive assessment frameworks, which detail the measurement rules and procedures that accreditation panels must apply. From the first (VSNU, 1987) to the current (NVAO, n.d.) edition of this framework, successive changes between editions reflect both contemporary currents in higher education policy (Weenink, 2025), but also the practical experiences of (thousands of) peer review panels.

The latter can be traced in several reports, which document the attempts by accreditation panels to operate the then-current assessment framework. One early framework, for example, is criticized by panels for being too detailed, leading not to more objectivity but to excessive bureaucracy (Goedegebuure et al., 2002). Conversely, an earlier framework was deemed by one panel to be too *un*specific, its criteria described as a blunt instrument (VSNU, 1996, pp. 3-4).

Grappling with HEQ measurement through social deliberation. Successive frameworks have attempted to address such measurement-related issues. Nevertheless, similar qualms are expressed by contemporary accreditation panellists in my interviews with them. Panellists continue to grapple with the way that HEQ is measured in the Dutch accreditation system: how it is defined, represented and proceduralised (Cartwright & Runhardt, 2014). Such qualms become especially manifest during an accreditation panel's social deliberation and judgment aggregation process. In the final part of my paper, I will give some examples of this phenomenon, and suggest some ways in panels use social deliberation and judgment aggregation to resolve their measurement-related issues.

References

Cartwright, N., & Runhardt, R. (2014). Measurement. In: N. Cartwright & E. Montuschi (Eds.), *Philosophy of Social Science: A New Introduction* (pp. 265–287), Oxford: Oxford University Press.

Chang, H. (2004). *Inventing Temperature: Measurement and Scientific Progress*. Oxford: Oxford University Press.

Goedegebuure, L., Jeliazkova, M., Pothof, F. & Weusthof, P. (2002). *Alle begin is moeilijk. Evaluatie van de proefaccreditering HBO.* Center for Higher Education Policy Studies.

Nederlands-Vlaamse Accreditatieorganisatie [NVAO]. (n.d.). *Beoordelingskader accreditatiestelsel hoger onderwijs Nederland*.

Rodriguez Duque, S., Tal, E., & Barbic, S. (2024). The role of ethical and social values in psychosocial measurement, *Measurement*, **225**, 113993.

Tal, E. (2013). Old and New Problems in Philosophy of Measurement. *Philosophy Compass*, **8.12**, 1159–1173.

Vereniging van Samenwerkende Nederlandse Universiteiten [VSNU]. (1987). *De Externe Kwaliteitszorg. Een gids voor de faculteiten ter voorbereiding op het bezoek van de visitatiecommissie.*

Vereniging van Samenwerkende Nederlandse Universiteiten [VSNU] (1996). *Onderwijsvisitatie Wijsbegeerte.*

Weenink, K. (2025). *Higher education quality and its contexts: How people make quality in interdependence.* Doctoral dissertation, Radboud University Nijmegen.

La Bête du Gévaudan, France,
allegedly killed more than honderd people in the period 1764--1767

Second Plenary Talk, Friday 09:00–10:00 h.

On Mathematics as a Language of Physics
Caspar Jacobs

"The Book of Nature is written in the language of mathematics”, so wrote Galileo. Indeed, contemporary physics is highly mathematical. But it is usually thought that such theories require an *interpretation* to make sense of them. We can think of an interpretation as a commentary on the theory’s formalism in an already-understood language. Recently, however, it has been alleged that mathematics can represent by itself. Theories already describe the world as being a certain way. Interpretation is not necessary. Moreover, this type of ‘*maths-first*’ representation supposedly lends itself naturally to an anti-metaphysical form of *structural realism*. In this talk I will not call into doubt the possibility of maths-first representation, yet I will try to show that maths-first representation is not necessarily structural.

In more detail, I will survey some options for the (meta)semantics of maths-first representation, none of which underwrite structuralism. I will also point out that such structuralism would require a form of incommensurability between mathematics and language. There is another approach which I favour: one on which mathematics and language are employed jointly to make sense of the physical world.

Galilei watching the heavens through a Dutch instrument, under a Van Goghish starry night sky

First Parallel Session, Friday 10:15–12:15 h.
Room 1

A New Counterfactual Theory of Actual Causation
Lennart Ackermans

When is an actual event caused by another actual event? Providing a definition of actual causation in a way that satisfies commonly held intuitions has long been thought to be an unsolved problem. However, much progress has been made recently. Gal- low (2021), for example, provides a new model-invariant counterfactual theory (*GC*). Andreas and Günther (2024, forthcoming) have presented a theory (*AGC*) that aligns with our intuitions surprisingly well. Yet each theory still faces notable challenges.

I provide a new theory of actual causation (NCC) that improves upon the current state-of-the art theories. The theory comes in two parts. The first part defines actual causation relative to a model, and the second defines what it means for a model to be appropriate for a real-world causal scenario. The two parts are combined to create a model-invariant theory of actual causation.

According to NCC, one determines whether an actual event A causes an actual event B (relative to a boolean model M) by first creating a baseline scenario in which exogenous variables apart from A may have a different value. Second, one intervenes upon all endogenous variables whose values have changed in the baseline scenario compared to the actual scenario (that is, their equations are removed from the model). Finally, A is said to cause B if B is counterfactually dependent on A in the post- intervention model from step 2. The key innovation of this definition is the second step, which eliminates inactive causal pathways in a simpler and more effective way than previous approaches.

In the second part, I introduce three new conditions for model appropriateness, called *exogenous deviancy*, *exogenous saturation*, and *endogenous saturation*. *Exogenous deviancy* and *exogenous saturation* draw on earlier work on normality (e.g., Halpern and Hitchcock, 2015; Hitchcock and Knobe, 2009). They require that all exogenous variables have deviant values and that no relevant variables with deviant values are excluded. *Endogenous saturation* requires that the internal part of the model is max- imally fine-grained, meaning that adding additional endogenous variables does not change certain aspects of the model's structure. A desirable feature of a theory of causation is that the choice of model does not affect its judgments. Given the above restrictions on model appropriateness as well as a number of standard restrictions, NCC's definition of causation, like GC, is locally model-invariant. This means that when variables are added and removed from a model (while retaining appropriateness), the model-relative definition makes the same causal judgments as before. I also show that AGC is not locally model-invariant, even given NCC's restrictions on model appropriateness.

Like AGC, NCC makes the intuitively correct causal judgments for known problematic scenarios. (In particular, it makes the same judgments for all examples given by Andreas and Günther in various papers, which include overdetermination, pree- mption, and threat-and-saviour scenarios.) In this respect, NCC is an improvement over all known theories of actual causation except for AGC. But it improves upon AGC by being simpler, model-invariant, and (arguably) by being counterfactual.

References

Andreas, H., & Günther, M. (2024). A lewisian regularity theory. *Philosophical Studies*, **181.9**, 2145–2176.

Andreas, H., & Günther, M. (forthcoming). Factual difference-making. *Australasian Philosophical Review*.

Gallow, J. D. (2021). A model-invariant theory of causation. *The Philosophical Review*, **130.1**, 45–96.

Halpern, J. Y., & Hitchcock, C. (2015). Graded causation and defaults. *The British Journal for the Philosophy of Science*, **66.2**, 413–457.

Hitchcock, C., & Knobe, J. (2009). Cause and norm. *The Journal of Philosophy*, **106.11**, 587–612.

Complex System: a Changing Knowledge Concept since the 18th Century

Maarten G. Kleinhans

The basic knowledge concept of complex system is central in ecology, earth science and climate science (Kleinhans 2025). Ladyman *et al.* (2013) define a *complex system* as an ensemble of many elements which are interacting in a disordered way, resulting in robust organisation and memory, where memory is defined as a robust, possibly hierarchical, order. Herewith, Ladyman *et al.* pursue a realist account of representation of patterns in the world by complex systems. However, their definition connects the concept with the concepts of interaction (and mechanism, and organization, and organism). Each of these are carriers of knowledge with several meanings that changed over time, so that defining one in terms of the other is obscuring the meaning. In this paper the concept and its articulation by users are traced from the late 18^{th} century to the early 21^{st} century.

In the past few years, the name of Alexander von Humboldt has been used by climate scientists as the first scientist to recognize complex systems and to position themselves on his shoulders. Humboldt's view was influential, for example Vernadsky cites it in his Biosphere system in the early twentieth century but denounces the organicist aspects. On the other hand, Bertalanffy, the system theorist, entirely ignores Humboldt. That contemporary scientists believe that Humboldt thought about complex systems in their terms is understandable but problematic. His detailed and often reproduced *Tableau* (*Naturgemälde*), produced as part of his 1807 essay on the geography of plants, depicts a mountain with vegetation zones in conjunction with observations of physical climate conditions. He used this graphic to represent the zonation of vegetation as related to climate zones on the entire planet. In *Kosmos* (1845), he clearly describes causal feedbacks. Humboldt's case is interesting because he does not call his view a system and explains why: the word was mainly used at the time for a priori imposed ordering principles, or, in other words, for systematic rather than systemic organization. On the other hand, he invoked physical forces as the ordering, or organizing, principles for observed vegetation patterns in nature, but he also invoked, obscurely, an organicist force of life while rejecting a version of vitalism. Humboldt politely but firmly rejects any association with systematic systems. I will show how definitions of complex system and of mechanism of the past decade changed while remaining entangled with each other and with the concepts of organization and organism, so that defining one in terms of the other is obscuring layers of changing meanings. Moreover, I will argue that the context and intention of use of the knowledge concept of complex system in practice is consistent with deflationary, rather than realist, accounts of representation.

References

Kleinhans, M. G. (2024). Causality and Complex Systems in the Geosciences. In: Illari, Ph and Russo, F. (eds). *The Routledge Handbook of Causality and Causal Methods*, pp. 652–659, Routledge.

Ladyman, J., Lambert, J., & Wiesner, K. (2013). What is a complex system? *European Journal for Philosophy of Science* **3**, 33–67.

Von Humboldt, A. und Bonpland, A. (1807) *Ideen zu einer Geographie der Pflanzen nebst einem*

Naturgemälde der Tropenländer, printed by F.G. Cotta, Tübingen.
Von Humboldt, A. (1845). *Kosmos. Einer physischen Weltbeschreibung*. Cotta'scher Verlag, Stuttgart und Tübingen, scanned in 2013 by the Deutsches Textarchiv.

Nomic truthlikeness in the light of a probabilistic representation of propositions

Theo A.F. Kuipers

In this paper philosophy of science, logic, and probability theory come together. There are already several general proposals for a quantitative measure for (actual and nomic) truthlikeness (e.g. Oddie, 1986; Niiniluoto, 1987; Zwart, 2001; Kuipers, 2023), all based on an underlying distance function between two propositional constituents or their substitutes. Such approaches are called similarity or likeness approaches, but we will call them more specifically 'horizontal' similarity approaches. Assuming a propositional language, we will focus on 'vertical' distance and similarity. Assuming two real-valued functions over the constituents, it is plausible to define the normalized distance between these functions in terms of the normalized sum of the absolute differences between the values, and hence the similarity as 1 minus this distance. We will show that this approach leads to at least two interesting distance measures between propositions, both of the content type (Zwart, 2001; Oddie and Cevolani, 2022). The one is the well-known, and much criticized, quantitative symmetric difference distance measure, the other is, as far as I know, a new, but plausible, probability-based, distance measure. We will start with the new one.

We first introduce a simple probabilistic representation of a proposition, guided by the *principle of indifference* and here called the corresponding propositional (probability) distribution, viz., for proposition X, with size $|X|$, i.e. the number of constituents making X true, assign them $1/|X|$ and 0 to all other constituents. We then determine the (normalized) distance between two propositions by applying the above indicated general definition of the (normalized) distance to the two corresponding distributions (Kuipers, 2024a). On the basis of the presupposed nomic truth, i.e. the strongest true proposition, characterizing the set of nomically possible constituents, we can now define the distance of an arbitrary proposition to the nomic truth, and hence its degree of nomic truthlikeness.

The probabilistic representation of a proposition and the indicated distance also enables several other interesting definitions: the strength of a proposition (relative to the tautology), the degree of inequality of a proposition, the degree of mutual dependence, or entanglement, as well as the degree of disorder of the atomic propositions within a proposition.

There are some other measures with which the probabilistic distance measure between propositions can be related or compared, all of the content type. First, there is a direct formal link with the so-called *fractional* distance measure between two quantities (Kuipers, 2024b), for which reason we call the present measure 'fractional distance related'. Second, since the fractional distance measure has a kind of twin measure, called the *proportional* distance measure, this suggests a variant of the probabilistic distance definition, called 'proportional distance related'. Third, and finally, it is plausible to compare these measures with the well-known 'symmetric difference' distance measure between propositions, the paradigm of a content definition, and here reconstructed as a 'vertical' definition, based on the Kronecker-delta function.

The three resulting truthlikeness measures will be illustrated and compared by a simple electric circuit, about which the nomic truth is easy to determine. More specifically, the three measures will be applied to three (nomically false) theories about the circuit. Finally, some issues for further research will be indicated, among which comparing and combining the new vertical measure with measures based on 'horizontal' similarity.

References

Kuipers, Theo, (2023). A coherent trio of, distance and size based, measures for nomic and actual truthlikeness, *Synthese*, **201**: 68.
Kuipers, Theo, (2024a), Truthlikeness and the Number of Planets, *Journal of Philosophical Logic* (2024) **53** :493–520.
Kuipers, Theo, (2024b), Degrees of Truthlikeness, Independence, Equality, and Order in Probabilistic Propositional Knowledge Representation, to appear in the *Journal of Logic and Computation,* online: https://doi.org/10.1093/logcom/exae031.
Niiniluoto, Ilkka, (1987). *Truthlikeness*, Reidel, Dordrecht.
Oddie, Graham, (1986) *Likeness to Truth*, Reidel, Dordrecht.
Oddie, Graham and Gustavo Cevolani, (2022), Truthlikeness, *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition), E.N. Zalta & Uri Nodelman (eds.),
Zwart, Sjoerd, (2001). *Refined Verisimilitude*. Dordrecht: Kluwer Academic Publishers.

First Parallel Session, Friday 10:15–12:15 h.
Room 2

Understanding and Controversy in Evolutionary Biology,
Ludo Schoenmakers

Understanding has become a focal concept in epistemology and philosophy of science. This holds true particularly for explanatory and objectual understanding. Here, explanatory understanding is the understanding 'why phenomenon p', where understanding p requires "a grasp of the reason why p" that comes in the form of an explanation (Hills 2016, 663). Objectual understanding, on the other hand, does not deal with particular phenomena, but covers a wider domain of reality. It "occurs when understanding grammatically is followed by an object, as in understanding the presidency, or the president, or politics, or the English language" (Kvanvig 2003, 191).

When it comes to explanatory and objectual *scientific* understanding, the two most developed accounts currently on offer are the contextual accounts of explanatory scientific understanding (de Regt 2017) and objectual scientific understanding (Elgin 2017). On de Regt's account, explanatory scientific understanding of phenomenon P obtains "if and only if there is an explanation of P that is based on an intelligible theory T and conforms to the basic epistemic values of empirical adequacy and internal consistency" (de Regt 2017, 92). Where the intelligibility of a theory is defined as "the value that scientists attribute to the cluster of qualities of a theory (in one or more of its representations) that facilitate the use of the theory" (de Regt 2017, 40). On Elgin's account, objectual scientific understanding of some domain D is achieved when a scientific account A of domain D is in reflective equilibrium, which means, amongst other things, that the various commitments of account A are consistent and co-tenable, that they exemplify the parts of reality that they are about in the right sort of way, and that account A has been formulated by a scientific community consisting of free and equal members.

Interestingly, so far, contextual accounts of scientific understanding have only been applied to scientific controversies in a limited sense. The account from (Elgin 2017) has not been applied to scientific controversies, while (de Regt 2017) provides an in-depth discussion of several controversial historical episodes in physics, but does not discuss scientific controversies as such. This is interesting because scientific controversies are often highly complex and heterogenous, and contextual accounts seems especially suitable for analysing these kinds of cases.

Perhaps the closest that (de Regt 2017) and (Elgin 2017) come to discussing scientific controversies is when they discuss the demarcation between science and pseudoscience. However, I argue that rather than focussing on demarcation, contextual accounts are better off focussing on philosophically more interesting and relevant cases of scientific controversy. Although a wide variety of definitions and descriptions of scientific controversies are on offer, in general, a scientific controversy can be described as a "publicly and persistently maintained dispute" between scientists about a belief or knowledge claim that "is held to be determinable by scientific means" (McMullin 1989, 51). While scientific controversies are sometimes described as exceptional and non-essential parts of science (cf. Dietrich 2020), their regular occurrence and their importance for the advancement of science have been widely recognized (cf. Engelhardt and Caplan 1989; Machamer, Pera, and Baltas 2000).

To illustrate how scientific understanding can be used to make sense of scientific controversy in evolutionary biology, I examine a particular controversy in the biological sciences, namely the controversy over the role of adaptationism in evolutionary biology as debated in the early nineteen eighties (including Gould and Lewontin 1979; Mayr 1983; Williams 1985). In this controversy, one side argues that evolutionary biology has become too adaptationist, and that this leads to an inferior kind of understanding of biological phenomena and the domain of evolutionary biology, while the other side denies this – or so I argue. I show that an important part of what fuels the scientific controversy over adaptationism are competing notions of scientific understanding of biological evolution.

In my understanding-based account of scientific controversy in evolutionary biology, I focus on how one side accuses another of inferior understanding, and how the other side defends itself. However, a more common approach to interpreting scientific controversies in evolutionary biology is to interpret these as relative significance controversies (Beatty 1997; Kovaka 2021; Deaven, forthcoming). In the final part of this paper, I argue that relative significance-based accounts of scientific controversy face problems with quantification and individuation. I argue that a qualitative notion of relative significance does potentially play a role in scientific controversies in evolutionary biology, but to explain the broader dynamics requires an understanding-based account, such as the one I have presented here.

References

Dietrich, Michael R. 2020. "What Is the Nature of Scientific Controversies in the Biological Sciences?" In *Philosophy of Science for Biologists*, edited by Kostas Kampourakis

and Tobias Uller, 235–54. Cambridge University Press.
Elgin, Catherine Z. 2017. *True Enough*. Cambridge, Massachusetts: The MIT Press.
Engelhardt, H.T., and A.L. Caplan, eds. 1989. *Scientific Controversies: Case Studies in the Resolution and Closure of Disputes in Science and Technology*. Cambridge, UK: Cambridge University Press.
Gould, S. J., and R. C. Lewontin. 1979. "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme." *Proceedings of the Royal Society of London. Series B, Biological Sciences* **205** (1161): 581–98.
Hills, Alison. 2016. "Understanding Why." *Noûs* **50.4**: 661–88.
Kvanvig, Jonathan L. 2003. *The Value of Knowledge and the Pursuit of Understanding*. Cambridge: Cambridge University Press,
Machamer, P.K., M. Pera, and A. Baltas, eds. 2000. *Scientific Controversies: Philosophical and Historical Perspectives*. Oxford University Press.
Mayr, Ernst. 1983. "How to Carry Out the Adaptationist Program?" *The American Naturalist* **121.3**: 324–34.
McMullin, Ernan. 1989. "Scientific Controversy and Its Termination." In *Scientific Controversies: Case Studies in the Resolution and Closure of Disputes in Science and Technology*, edited by H.T. Engelhardt and A.L. Caplan, 49–91. Cambridge, UK: Cambridge University Press.
Regt, Henk W. de. 2017. *Understanding Scientific Understanding*. Oxford University Press.
Williams, George C. 1985. "A Defense of Reductionism in Evolutionary Biology." In: *Oxford Surveys in Evolutionary Biology*, Richard Dawkins and Mark Ridley (eds.), **2**:1– 27, Oxford: Oxford University Press.

Clustering Models in Psychiatry: Practical Tools or Path to Natural Kinds?

Anna van Oosterzee

The recent advancements in artificial intelligence have made a significant impact across scientific disciplines, but its potential role in psychiatry and mental health care at large raises critical questions. In this paper, we consider the use of clustering models in psychiatric diagnostics, a type of unsupervised learning used to uncover subclassifications within psychiatric disorders, with the hope that they could discover natural kinds (Watson, 2023). We argue that it remains highly doubtful that these models can meet such expectations in psychiatry.

For over five decades, psychiatry has relied on the Diagnostic and Statistical Manual of Mental Disorders (DSM) to classify mental health conditions, but there is widespread agreement that the DSM does not contain valid categories of mental disorders. The current diagnostic system often results in highly heterogeneous groups, with significant overlap between conditions (comorbidity) and a failure to account for many patients' mental health problems. These limitations impede both scientific research and treatment effectiveness (Tsou, 2016).

In light of these challenges, many researchers are advocating for precision psychiatry — a more personalized approach aimed at tailoring treatment to individual patients. One promising approach within this paradigm is the use of clustering models, which aim to discover meaningful subgroups of disorders from large datasets without prior knowledge of categories. The *hope* is that these models will reveal previously unknown, therapeutically relevant categories of mental health problems, offering new avenues for treatment. However, we argue that while clustering models are powerful tools, they are unlikely to uncover "natural kinds" in psychiatry due to the inherent limitations of data selection, feature extraction, and model parameters.

Clustering models depend on the choice of relevant variables or features and quality of the data collected.In psychiatric research, data is often shaped according to existing research assumptions. Decisions such as which patient groups to analyze, which features to extract and how many clusters to select are shaped by practical, and often political, considerations rather than purely scientific ones.

Instead, the outcomes reflect the specific questions researchers aim to answer, often resulting in what philosopher Peter Zachar (2003) and others call 'practical kinds' — categories that are useful for particular purposes but do not necessarily reflect the underlying nature of mental disorders.

Through an examination of relevant case studies, such as Zhang et al.'s (2021) use of sparse clustering to identify subtypes of Major Depressive Disorder (MDD) and Post- Traumatic Stress Disorder (PTSD), we highlight how clustering models produce results that are tailored to specific clinical and research goals. These models may indeed identify clinically relevant subtypes, but they are unlikely to reveal robust ontological truths about mental disorders.

In conclusion, while clustering models hold promise for improving psychiatric classification and treatment, they are best understood as tools that generate useful, context-dependent categories rather than as methods for uncovering natural kinds. It is crucial that we recognize the limitations of these models and focus on how they can be applied effectively in psychiatric practice, rather than expecting them to transform our understanding of mental health at a deeper level.

Disorders (DSM-5-TR). American Psychiatric Association Publishing.
Tsou, J. Y. (2016). Natural kinds, psychiatric classification and the history of the DSM. *History of Psychiatry*, **27.4**, 406–424.
Watson, D. S. (2023). On the Philosophy of Unsupervised Learning. *Philosophy & Technology*, **36.2**, 28.
Zachar, P. (2003). The practical kinds model as a pragmatist theory of classification. *Philosophy, Psychology and Psychiatry*, **9.9**, 219–227.
Zhang, Y., *et al.* (2021). Identification of psychiatric disorder subtypes from functional connectivity patterns in resting-state electroencephalography. *Nature Biomedical Engineering*, **5.4**, 309–323.

Microbiomedical Research and the Myth of Ethno-racial Categories
Aline Potiron

The increasing use of ethno-racial categories in microbiome studies has sparked debate among scholars, with concerns surrounding the problematic narrative of "Westernized" vs. "non-Westernized" that reifies racist views (for a review, see Rawson 2024). I focus on the more recent trend in microbiome studies, roughly after 2018, where ethno-racial categories are used to study human health disparities to inform personalized medicine. I argue that this use is problematic and should be abandoned.

In this paper, I question the inferential power of ethno-racial categories in microbiome studies in the context of personalized medicine on three points.

Firstly, in the context of medical genetics, it has been argued that the definitions of ethno- racial categories are ambiguous and inconsistent, creating epistemic uncertainty (Malinowska and Serpico, 2023). I argue that this conceptual imprecision also plagues microbiome studies. These studies often use interchangeably different concepts such as ‘race’, ‘ethnicity’, ‘geography’, and ‘nationality’. Moreover, few studies define these terms before using them or do not use them consistently throughout one paper. Additionally, when institutional or more general definitions backed up such papers, those definitions are not consistent globally and depend mainly on the location of the researchers (e.g., USA vs Netherlands). Ultimately, this situation prevents comparative analyses of different studies and hinders scientific communication.

Secondly, these categories reduce multiple dimensions, including human genetic diversity, diet, and socio-cultural dimensions, into one category, which can result in stereotyping and essentializing ethno-racial categories through the microbiome. This unclarity of the actual variables of interest also creates confusion in cause-effect relationships: It is unclear whether the ethno-racial category is a cause of the microbiome composition or if it is an effect of the diet (or another dimension) mediated by the microbiome.

Finally, few studies have found consistent, reliable associations between microbial taxa and disease or health phenotypes. Similarly, no taxa are reliably and consistently associated with a particular human allele. Researchers tiptoed on what is really relevant for a disease in the ethno-racial category (i.e., a genetic variant, a nutrient in the diet, etc.). However, these factors could lead to different classifications of the same human population, rendering ethno- racial categories useless for treating a particular individual.

Overall, I argue that ethno-racial categories in microbiome studies have limited inferential power. The risks of using ethno-racial categories in microbiome studies, including poor scientific communication, attributing a spurious causal role to these categories, stereotyping, and misplacing individuals into categories that may not be relevant for treatment or risk- factor assessment, outweigh any potential benefits. Instead, researchers should focus on the actual variables they are interested in, such as diet, to avoid racist descriptions and the micro-biomization of ethno-racial categories.

References

Malinowska, J. K. and Serpico, D (2023). Epistemological Pitfalls in the Proxy Theory of Race: The Case of Genomics-Based Medicine. *The British Journal for the Philosophy of Science*, to appear: https://doi.org/10.1086/727957

Rawson, A. J. (2024). Anti-racism, racism, and the microbiome: A review. *Progress in Environmental Geography*, **3.2**, 137–159.

Lupine dinner party

Second Parallel Session Friday, 13:15 – 15:15 h.
Room 1

The Loss of Ideology in the Philosophy of Science
Thijs Ringelberg

Social epistemologists of science often use an economic approach (EA) (Desmond 2021) (e.g. Kitcher 1990; Strevens 2003; Heesen and Bright 2021). I provide an account of the history of the EA to argue that its rise coincides with a decline in interest in ideological aspects of the Philosophy of Science (PoS). I trace this loss of ideology to influences from sociology and economics, and argue that it creates problems that require remedy.

I use "ideology" in a non-Marxist sense, following Bell (1962), to denote action-oriented systems of beliefs. PoS is ideological if it offers scientists behavioural norms and values ('ethos') aligned with a broader narrative about science's purposes and functioning. Although the EA does offer normative analyses, these typically focus on the 'material base' of science (e.g. incentive structures, communication networks), while neglecting ethos. EA-analyses thus tend not to engage with ideological aspects of science.

I trace this lack of engagement with ideology to the EA's origins. Its foundations were laid by Philip Kitcher (1993) in response to challenges from the Sociology of Scientific Knowledge (SSK). PoS had traditionally maintained that social and causal factors could only explain the beliefs of scientists in instances where scientists had been led astray; explanations of belief in rational theory, however, required reconstructions of the internal justificatory structure ('internal history') of the theory. Against this, SSK advocated the symmetry thesis, arguing that all beliefs could be explained exclusively by sociological factors (Bloor 1991). SSK's arguments for the symmetry thesis involved rejecting the universalist aspirations of rationalist terms like "rational" or "knowledge."

In response, Kitcher carved out a new role for PoS. He accepted symmetry while holding on to universalism by emphasising the possibility of scientific progress. A distinction could be made, he argued, between social structures that caused progress and structures that caused decline. The role of PoS, then, was to tell these two kinds of social structures apart. He articulated this new role by borrowing from microeconomics to build models of scientific social structures. I argue that embedded in these modelling techniques came an implicit methodological commitment to anti-paternalism: policy – and therefore the theoretical analyses on which policy is based – should be effective regardless of individuals' normative convictions.

Thus, on my account of its history, the EA has incorporated two anti-ideological ideas. From SSK (via the symmetry thesis), it inherited the descriptive claim that scientific beliefs can be explained by the material base of science rather than scientists' normative convictions. From

economics (via anti-paternalism), it inherited the normative injunction to abstract away from scientists' normative convictions in its analyses.

The resultant blindness to ideology, I argue, is problematic: it causes the EA to underestimate the role of ideology in the functioning of science as well as science policy. I therefore argue that the EA should regain a view of ideology, and suggest that more deliberative approaches to PoS, such as those developed by (Longino 1990), serve as inspiration for successfully integrating the symmetry thesis whilst retaining an ideological function.

References

Bell, Daniel. 1962. *The End of Ideology: On the Exhaustion of Political Ideas in the Fifties*. Cambridge, Massachusetts: Harvard University Press.

Bloor, David. 1991. *Knowledge and Social Imagery*. 2nd ed. University of Chicago Press.

Desmond, Hugh. 2021. Incentivizing Replication Is Insufficient to Safeguard Default Trust, *Philosophy of Science* **88.5**: 906–17.

Heesen, Remco, and Liam Kofi Bright. 2021. Is Peer Review a Good Idea? *The British Journal for the Philosophy of Science* **72.3**: 635–63.

Kitcher, Philip. 1990. The Division of Cognitive Labor, *The Journal of Philosophy* **87.1**: 5–22.

Kitcher, Philip. 1993. *The Advancement of Science: Science without Legend, Objectivity without Illusions*. New York; Oxford: Oxford University Press.

Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, New Jersey: Princeton University Press.

Strevens, Michael. 2003. The Role of the Priority Rule in Science, *Journal of Philosophy* **100.2**, 55–79.

Revisiting the Analogy between Grounding and Causation
Martin Voggenauer

Grounding is often regarded as a kind of metaphysical causation (Wilson 2018). Some even consider causation as a particular grounding (or building) relation (Bennett 2017). However, although there are strong similarities in content, structure, and form, the analogy is not uncontroversial given some notable differences (Bernstein 2016). In this talk, I argue for a restricted analogy that does justice to both the striking similarities and the key differences. Specifically, I argue that the analogy holds only between causation and particular grounding relations, namely those that connect events on different levels of nature (or facts about such events).

On the one hand, grounding shares some overwhelming content-related, structural, and formal similarities with the concept of causation, as in particular Jonathan Schaffer (2016) and Alastair Wilson (2018) argue. Specifically, both concepts connect prior and subsequent entities and their events (or facts) in the most general way. Furthermore, they both back explanations. Eventually, both concepts are irreflexive, asymmetric, and transitive (or at least acyclic) relations and can therefore be formalized as directed acyclic graphs.

On the other hand, Sara Bernstein (2016) has put forward some notable differences and concluded that the analogy is not well motivated and should be rejected. First, grounding is a very broad concept that is often considered as primitive and rather intuitive. By contrast, causation has always played a prominent role in science and everyday thinking and has undergone a multitude of philosophical attempts at analysis. Moreover, there are several special features of causation that grounding does not seem to share. For example, we can make sense of productive, negative, and probabilistic causation, but apparently not of productive, negative, or probabilistic grounding.

In this talk, I argue that the best way to counter Bernstein's objections while accounting for the striking similarities between grounding and causation is to restrict the analogy. To this end, I follow Jessica Wilson's (2014) distinction between a unified 'big-G' grounding concept and various specific 'small-g' grounding concepts. Building on this distinction, I argue that the grounding relations that are analogous to causation are those that connect events on different levels of nature (as opposed to relata on different levels of metaphysics, see, e.g., McKenzie (2022)).

I show that the restricted analogy between grounding and causation fares way better than the common grounding-causation analogies. First, we get further immediate similarities. Both causation and event grounding are worldly relations between events. While causation connects events over time, this kind of grounding connects events over levels of nature. Second, it helps to overcome Bernstein's (2016) most serious objections, while maintaining the content- related, structural and formal similarities.

In detail, the restricted analogy immediately overcomes the difference that there is a distinction between a unified big-G relation and particular small-g relations for grounding, but no analogous distinction for causation. Moreover, the difference that (big-G) grounding is usually considered a primitive and arguably unintuitive concept is overcome, since the relation between events on different levels of nature is a way more intuitive concept that can be analyzed by recourse to concepts such as constitution, composition, or realization, etc. In addition to these general differences, there are some specific structural differences, such as missing grounding analogs for productive, indeterministic, and negative causation, that can be overcome by the restricted analogy as well – or so I argue.

References

Bennett, Karen (2017): *Making Things Up.* Cambridge: Cambridge University Press.
Bernstein, Sara (2016): Grounding is not Causation, *Philosophical Perspectives,* **30**, 21–38.
McKenzie, Kerry (2022): *Fundamentality and Grounding.* Cambridge, UK: Cambridge University Press.
Schaffer, Jonathan (2016): Grounding in the Image of Causation. *Philosophical Studies,* **173**, 49-100.
Wilson, Alastair (2018): Metaphysical Causation, *Noûs,* **52**, 723–751.
Wilson, Jessica (2014): No Work for a Theory of Grounding. *Inquiry,* **57**, 535–579.

What can we learn from replication efforts about scientific pluralism?
Stephanie Meirmans

Recently, the Dutch funding organization NWO issued a call for doing *replication studies*. Researchers in the social sciences, in the medical sciences, and in the humanities could apply to this call. Here, I argue that this call can be seen as an epistemic science policy experiment. It operationalized an epistemic framework of diagnostics replication. This framework entails a science done in a 'Popperian' way, as error correction, as a way of self-correction in science.

Several philosophers had warned, however, that this kind of epistemic framework does not provide an epistemic fit across all disciplines, or 'ways of knowing'. For example, Bart Penders and colleagues have pointed out that the humanities might not fit into this epistemic mold, and that it would be dangerous even to attempt forcing humanities scholars to aim towards reproducibility. Also Sabina Leonelli pointed out that there are different kinds of reproducibility across different scien9fic disciplines. It is precisely these kinds of potential tensions due to a scien9fic pluralism that make it so interesting to empirically track what happened when the underlying epistemic idea of this funding call was 'released into the wild' (i.e. into scientific practice).

In this talk, I will explore more in-depth what was the local context for issuing this call at all, and what specific kind of epistemic framework it entailed. I will also describe and analyse what happened in the funded projects: what the motives were to apply at all, what the surprises, the expecta9ons, the problems, the outcomes and the impact of the projects. For this, I use ethnographic material that I have gathered since 2021. This includes interviews with several of the main people involved in se]ng up the funding call, with funding officers, with researchers who got the funding and did the replication, with involved students and other researchers in these replication efforts, and with original authors of the studies being replicated. It includes observations of experiments, of handling data and texts in practice, and of discussions in workshops and conferences about the replication work. It also includes gathering material such as published papers, paper drafts, PowerPoint presentations and proposals. For this talk, I use material for eight NWO-funded replication projects (across all three disciplines). I also add ethnographic material gathered in six additional humanities projects and one medical study, all doing replications, and all performed in the Netherlands around the same time.

Perhaps unsurprisingly, I found that researchers across different fields started their endeavor with very different assumptions, hopes, and goals. Much more surprising was that in the end all researchers, across disciplines, ran into or pointed out very similar issues and problems with the replica9on work. Many also experienced added value in doing replication work in a very similar manner. In this talk, I will outline and highlight a number of such rather surprising similarities across fields; similarities that might perhaps not have been expected from a traditional framework of different scientific styles across disciplines.

Second Parallel Session Friday, 13:15 – 15:15 h.
Room 2

The Scope and Resolution of Neyman's Paradox
Samuel C. Fletcher

Neyman's paradox concerns the objectivity of statistical testing that considers only the hypothesis tested and not any alternative to it. Neyman's target was one of Ronald Fisher's approaches to testing (e.g., in his 1950). According to this approach, before collecting data X, one selects a test statistic $T(X)$ of the data whose probability distribution can be determined or estimated assuming the truth of the hypothesis H under test. If T is a continuous (discrete) random variable, then one strictly and totally orders its values according to their decreasing probability density (probability mass) f_H under this hypothesis, viz., for values of the statistic t_0 and t_1, $t_0 <_H t_1$ if and only if $f_H(t_0) > f_H(t_1)$. One interprets this ordering as a qualitative relation of extremeness with respect to the mode. Finally, once the data is collected, one then calculates the p-value value of the statistic evaluated at the data, t, under the hypothesis H under test: $P_H(T \geq_H t)$. The p-value quantifies the extremeness of the data, which thus provide evidence against H to the extent to which the p-value is small.

While admitting that there have been many scientifically fruitful applications of the approach of Fisher's, Neyman asked how it could possibly adjudicate between two different tests with conflicting outcomes. He considered two different statistics T_1 and T_2 of the same data such that: (i) under the hypothesis under test, T_1 and T_2 have the same probability density functions, and (ii) to the extent that T_1 yields a low p-value, T_2 yields a high p-value, and vice versa. On account of (i), one cannot prefer one of the statistics over the other by its distributional properties. On account of (ii), Fisher's approach leads to conflicting results about the evidence against the hypothesis under test, depending on the seemingly arbitrary choice between T_1 and T_2. In particular, if one adopts a threshold p-value for rejecting the hypothesis, then employing T_1 leads to rejection only if T_2 does not, and vice versa. The paradox is that Fisher's seemingly reasonable procedure for statistical testing yields inconsistent or arbitrary results.

Neyman understood the existence of paradoxical counterparts to arise only for non-comparative accounts of statistical testing, ones whose tests depend at most on the hypothesis under test and no other. For, he averred, one could dissolve the paradox by taking into account hypotheses alternative to the one under test. With

Egon Pearson, he proved the famed Neyman-Pearson lemma, which states roughly that if one is testing one simple statistical hypothesis against another, then the likelihood ratio of these hypotheses is the unique statistic (up to a positive constant of proportionality) that maximizes the probability of finding evidence against the one if it is false.

I propose that one can dissolve the paradox by demanding invariance under transformations that preserve the extremeness ordering of the data, because this ordering is part of the data's evidential content. Although imposing ordering requires attributing more structure to test statistics than Neyman, I contend that it is nevertheless essential to the function and interpretation of significance tests. Yet that structure must draw on how some ways of grouping data distinguish between some hypotheses to the exclusion of others. Thus, my dissolution actually vindicates his core insight that if statistical testing is to be coherent, it must be conducted on a comparative basis.

Decision Models and Performativity Problems
Oyku Ulusoy

Decision theory provides formal frameworks for modelling and assessing the expected gains of one's choices (Savage, 1954; Steele & Stefánsson, 2020). However, can decision-makers effectively employ decision theory to rationally assess their alternatives when confronted with significant changes to their future selves and preferences? I argue that there exists a distinct type of choice problem in which the way we frame a choice-setting within a modelling framework directly shapes our conclusions about the rational course of action. I study this issue as a performativity problem for decision theory.

I conceptualize decisions that shape our future lives as transformative choices, following Paul (2014; 2015). Paul asserts that individuals cannot be adequately informed or advised about how their lives will be affected by new phenomena until they directly experience them. This lack of prior experience hinders their ability to form well-informed evaluations of the potential outcomes associated with these phenomena. She argues that, because individuals cannot fully comprehend how their experiences will alter their deliberations, they are unable to make rational choices between alternative standpoints and preferences.

Paul further contends that decisions leading to new discoveries and preference changes defy resolution through standard decision-theoretic approaches. Nevertheless, decision theory offers a range of model-theoretic frameworks that enable comparative analysis. This underscores the importance of studying methodological perspectives for modelling transformative choices. Accordingly, I identify two model-centric problems in applying decision theory to such cases.

The first problem, as Titelbaum (2013) argues, is that while we rely on model-derived observations to study choice phenomena, models necessarily omit certain characteristics of the phenomena they aim to represent. This limitation is an inherent drawback of model-based studies of rational choice. In the context of transformative choices, I argue that the challenge lies in the fact that outcomes in decision models are typically described based on pre-existing preferences. However, in transformative cases, new preferences are formed retrospectively, making any decision that leads to those future preferences appear rational in hindsight. This creates an epistemic performativity problem (Basshuysen, 2023), where decision models can rationalize any outcome of transformative choices as optimal.

This raises a related second question: how do we assess the suitability of models when studying transformative choice phenomena? This question can be framed as a practical performativity problem: which preferences are worth cultivating when facing fundamental changes in the organization of one's life? Addressing this issue requires the study of additional norms to evaluate the suitability of model frameworks for their target domains, which may vary depending on the specific objectives of the models (Broome, 2013). I argue that different modelling approaches may generate competing choice strategies, leading to distinct optimal actions with significantly different consequences. In addressing how to evaluate model-centric rationality norms, I propose examining whether a particular transformative decision can be deemed rational relative to the decision-maker's own understanding within different modelling frameworks, such as those proposed by Steele & Stefánsson (2021). This approach allows us to analyse whether models are appropriate for their intended targets by focusing on the claims they enable us to make about which strategies, in the face of contemporary uncertainties, appear more desirable (Beck & Jahn, 2021; Roussos, 2022; forthcoming).

References

Basshuysen, P. van. (2023). Austinian model evaluation. *Philosophy of Science*, **90.5**, 1459–1468.

Beck, L., & Jahn, M. (2021). Normative Models and Their Success. *Philosophy of the Social Sciences*, **51.2**, 123–150.
Broome, J. (2013). *Rationality Through Reasoning*. Wiley-Blackwell.
Paul, L.A. (2014). *Transformative Experience*. Oxford University Press.
Paul, L.A. (2015). What you can't expect when you're expecting. *Res Philosophica*, **92.2**, 149–170.
Roussos, J. (2022). Modelling in Normative Ethics. *Ethical Theory and Moral Practice*, **5**:1–25.
Roussos, J. (forthcoming). Normative Formal Epistemology as Modelling. *The British Journal for the Philosophy of Science*.
Savage, L. (1954). *The Foundations of Statistics.* Dover Publications.
Steele, K. & Stefánsson, H.O. (2020). *Decision Theory*. In: Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition).
Steele, K., & Stefánsson, H.O. (2021). *Beyond Uncertainty: Reasoning with Unknown Possibilities*. Cambridge, UK: Cambridge University Press.

Measuring Psychological Attributes in the Network Framework
Riet van Bork

In psychology, measurement of mental attributes relies on statistical models: so-called 'measurement models'. Existing measurement models are developed within a particular framework in which the attributes are assumed to be latent common causes of observed behaviors. However, more recently, 'network theories' of psychological attributes have gained traction in psychology, in which the assumption that the attribute acts as a common cause of the observed behaviors is explicitly rejected. Instead, behaviors, cognitions and feelings interact directly with each other, forming a complex system that can be represented as a network.

Although the development of psychological network theory has led to many new analytic techniques, it has not yet resulted in a psychometric framework that can guide measurement. For example, claims such as that some people are *more* extraverted than others, or that a particular school intervention has *increased* a person's ability, require that different levels of these psychological attributes can be compared across people or time. While the traditional framework of latent variable models accommodates such comparisons, it is unclear how to make such comparisons from networks of these constructs. In this talk, I will discuss some possible first steps in developing measurement theory for psychological attributes that are studied within a network theory, and the challenges involved.

One of these first steps is to conceptualize the psychological attribute that is targeted for measurement as a property of the network. For example, the psychological attribute could be conceptualized as the mean activity of all nodes in a large network. To take the construct 'extraversion' as an example, this would mean that there is a large network of behaviors that cluster together and that we associate with extraversion, and that a person's level of extraversion is conceptualized as the proportion of nodes in this network that are active (i.e., the proportion of these behaviors that this person engages in). Similarly, a person's level of depression severity could be conceptualized as the proportion of symptoms in a large depression network from which the person suffers. If this property of the network is chosen as the measurand, a next step is to develop a statistical model that describes the relation between this property and observable item scores. For example, a possible assumption is that when a test is administered, only a small random sample of nodes from the total network is observed. Also, one can make assumptions about the presence or absence of measurement error in observing these behaviors or symptoms. In describing these steps of formulating a measurement model for network properties, I will draw connections to existing psychometric theories.

The question of how network theories of psychological properties change the way we measure those properties is not only relevant in the context of psychometrics, but also provides an interesting case to look at how changes in theory can inform measurement. I therefore plan to consider Chang's (2004) notion of 'epistemic iteration', which describes how theory about a concept and measurement practices for that concept iteratively revise each other, in this context.

Wolverine,
not spotted on the Veluwe
so far

Third Parallel Session Friday, 15:30–17:30 h. Room 1

Broke But Not Worth Fixing? On Adjusting Effect Sizes for Research Biases

Ina Jäntgen

To inform clinical decision-making, biomedical researchers measure the effect sizes of tested treatments, and such effect sizes are often amalgamated in meta-analyses. However, these effect sizes can be, and commonly are, biased. Studies reporting effect sizes can be subject to *methodological biases* that threaten their internal validity, such as confounding. Moreover, a body of evidence can be subject to *meta-research biases*, such as publication bias, which affect amalgamated effect sizes. In response to such biases, guidelines suggest adjusting effect sizes whenever possible before reporting them to decision-makers (Higgins et al. 2023). Likewise, philosophers have proposed models for adjusting for meta-research biases (Erasmus 2023)

But when does adjusting biased effect sizes help inform clinical decision-making? An intuitive answer is always. After all, a less biased effect size conveys more accurate information about a treatment's effectiveness. And the more accurate the information we have, the better our decision- making, or so the thought. In this paper, I argue that this intuitive answer is mistaken from the perspective of rational decision-making. Sometimes, an adjusted and a biased effect size provide, in expectation, equally valuable information to choose between treatments rationally, and sometimes, the expected additional value of learning an adjusted rather than a biased effect size is not worth any costs of adjusting.

To establish these results, I first model a decision between treatments tested in trials using expected utility theory (c.f. Sprenger and Stegenga 2017; Jäntgen 2023), focusing on effect size measures for binary outcome variables, such as the risk difference. Relying on the decision-theoretic concept of the *expected value of information* (Good 1967), I ask: when is the expected value of learning an adjusted effect size higher than the expected value of learning a biased effect size for an agent rationally deciding between treatments? I show that sometimes learning an adjusted or a biased effect size is, in expectation, equally valuable, and sometimes learning an adjusted effect size has only marginally higher expected value. These results hold for several methodological and meta-research biases and both absolute and relative effect size measures. I conjecture that whether adjusting has

a higher expected value than not adjusting depends on (a) the type of research bias, (b) the effect size, and (c) the method of amalgamation. More research is needed to investigate precisely when adjusting matters. Finally, I criticise Erasmus' (2023) model for adjusting effect sizes since it demands adjusting when doing so is not expected to be valuable. Overall, learning adjusted effect sizes instead of biased ones is not always better for rational decision-making.
This result may have practical bite; adjusting for research biases requires gathering and processing data – often a laborious process. My results show that such costs of adjusting may not always be worth it when it comes to informing decision-making. This observation sets the ground for discussing when researchers should adjust effect sizes to inform clinical decision-making and when they may not. Overall, this paper adds neglected but important decision-theoretic considerations to recent debates on adjusting biased effect sizes in biomedical research.

References

Erasmus, Adrian. 2023. The Bias Dynamics Model: Correcting for Meta-Biases in Therapeutic Prediction. *Philosophy of Science*, April, 1–10.

Good, I. J. 1967. On the Principle of Total Evidence. *The British Journal for the Philosophy of Science* **17.4**: 319–21.

Higgins, JPT, et al., eds. 2023. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.4 (Updated August 2023). Cochrane.

Jäntgen, Ina. 2023. How to Measure Effect Sizes for Rational Decision-Making. *Philosophy of Science*, 1–17.

Sprenger, Jan, and Jacob Stegenga. 2017. Three Arguments for Absolute Outcome Measures. *Philosophy of Science* **84.5**, 840–852.

HARKing and Meta-Analysis: A Novel Perspective
Boris Kuiper

Meta-analyses are an increasingly prominent form of research synthesis and research integration where the results of multiple studies are combined using statistical methods. They are often regarded as the best and most objective evidence available about such vital matters as the efficacy of a medical intervention. There is a lively philosophical discussion about the objectivity and reliability of meta-analyses, and whether their status as the 'platinum' standard of evidence is warranted. This talk is about one aspect of this debate: whether their retrospective nature undermines their confirmatory value.

That meta-analyses are retrospective is hard to argue against: they are (almost always) written when the results of the primary research are publicly available. Furthermore, experience teaches us that it, while it is possible to preregister a meta-analysis, it is near-impossible to follow the preregistered analysis exactly. Many choices, such as those about the exact research question, inclusion criteria, and details of the analysis, may have to be revisited and amended during the analysis. Consequently, meta-analyses constitute a form of hypothesizing after results are known ('HARKing').

HARKing is generally frowned upon by most scientists, with several standard accounts for why it undermines the validity of results. Simultaneously, as it is inescapable in meta-analyses, we cannot just disavow it entirely; we need a good account of exactly why – and when – it is problematic. Aydin Mohseni, in a recent article, convincingly shows that these standard accounts fail to show that HARKing undermines the confirmatory value of results. Does this mean that HARKing is fine in meta-analyses?

Not quite, or so I argue. Mohseni makes two assumptions which I argue are unrealistic of both meta-analyses and scientific inference in general. First, he makes no distinction between a scientific hypothesis and the statistical analysis by which it is tested. The second is that he considers only two reporting protocols: a researcher either reports randomly among significant results, or she reports the most significant results. Following Gelman and Loken, I argue that there are generally many statistical analyses possible for the same scientific hypothesis, especially in meta-analyses. Given enough such analyses it becomes more and more likely that at least some of them will yield significant results.

So, what does that mean for meta-analyses? Given such a one-to-many mapping from statistical analyses to scientific hypotheses, the mechanism by which the researchers choose one (or perhaps a small number) of them to report becomes crucial. In the worst case, a researcher that will always choose a statistical analysis supporting a particular favoured hypothesis, when enough independent analyses are available, will always find and report significant results for that hypothesis. Perhaps good news for the researcher, but under such conditions no confirmatory value whatsoever is bestowed upon that hypothesis, regardless of results.

What is perhaps more worrying than the possibility of such biased researchers is that the confirmatory value of the results now depends on the reporting mechanism of the researchers writing the analysis – something that is not discernible from reading the finished report.

This constitutes a novel account of why HARKing can be dangerous that can apply beyond meta- analyses. It serves to highlight the importance of subjective factors going into statistical analyses. Furthermore, it yields important practical recommendations about the interpretation of different parts of meta-analyses and about certain statistical practices regarding them.

Science, Values, and the Interspecies Standpoint
Claudia Cristalli

Values play an important role in every scientific endeavour, yet this recognition alone is not enough to enough to know what to do with them. The issue is particularly compelling in ecology, where philosophers struggle between the option of defining concepts like "biodiversity" intrinsically valuable, or valuable because humans can get some potential benefits from it – such as the possible medicinal properties that Western science may still extract from richly biodiverse forests (Justus 2021; Colyvan et al. 2009; Faith 1994). Cost-benefits analysis and decision theory may be useful to make decisions once we know what our values are, but it does not help to reflect about the values that our science is currently using, nor about the implications for our future values of our current scientific practices (Ratti & Russo 2024, 8; Elliot 2017). Feminist scholars have long posed the question, "Whose values?," and standpoint epistemologists have emphasized the situatedness of knowledge previously deemed universal.

In this paper, I argue that the most responsible answer to such challenges for both scientists and philosophers of ecology is to expand the scope of who is entitled to contribute concepts and values to our scientific practices. Drawing from two recent studies on human-animal interaction, I propose that nonhuman animals exhibit their values and knowledge of the environment in interaction with other fellow animals, including humans. If this proves correct, then studying these interactions is key to understanding the nonhuman animal's standpoint. Moreover, if theories and methods have value-laden implications, the shift of attention to human-animal interactions should facilitate our appreciation of nonhuman animals' standpoints.

The first case study concerns the ongoing collaboration between human honey-hunters and bird honey-guides (Spottiswoode & Wood 2023; van der Wal et al.

2022; Cram, van der Wal, et al. 2022). The second focuses on shepherds in Southern France, who went back to the discontinued practices of transhumance after farmed pastures became uneconomical (Despret & Meuret 2015, 20). In spite of the differences between these two cases, show human and non-human animals acquire knowledge and benefit from their interaction. In both instances, the animals' knowledge is crucial for shaping the environment in which both humans and animals live. By focusing on the human-animal relationship, I makes the animal perspective accessible to philosophical inquiry, and make their presence visible as a standpoint. Animals bring certain interests to their environment and specialized knowledge of it; humans interested in studying this environment would do well to take note.

References

Colyvan et al. 2009. Philosophical Issues in Ecology: Recent Trends and Future Directions. *Ecology and Society* **14.2**: 22.

Cram, van der Wal, et al. 2022. The ecology and evolution of human-wildlife cooperation. *People and Nature* **4**, 841–855.

Despret & Meuret 2015. *Composer avec les moutons. Lorsque des brebis apprennent à leurs bergers à leur apprendre.* Cardère éditeur.

Elliott 2017. *A Tapestry of Values: An Introduction to Values in Science*, Oxford: Oxford University Press.

Faith, D.P. 1994. Phylogenetic pattern and the quantification of organismal biodiversity. *Philosophical Transactions of the Royal Society: Biological Sciences* **345**: 45–58.

Justus 2021. *The philosophy of ecology: an introduction*. Cambridge, UK: Cambridge University Press.

Ratti & Russo 2004. Science and values: a two-way direction. *European Journal for Philosophy of Science* (2024) **14**:6.

Spottiswoode & Wood 2023. Culturally determined interspecies communication between humans and honeyguides. *Science* **382**, 1155–1158.

Wal, van der, *et al.* 2022. Awer Honey-Hunting Culture With Greater Honeyguides in Coastal Kenya. *Frontiers in Conservation Science*, **2**, 1–7.

Third Parallel Session Friday, 15:30–17:30 h.
Room 2

Medical research without Big Pharma: It's preferable, it's profitable, and it's practicable
Hans Radder

This paper addresses the patent practices for prescription drugs by big pharmaceutical companies. We argue that medical research without such patents is epistemically, socially and morally preferable, economically and financially profitable, and socio-politically and organizationally practicable. Along the way, we emphasize the importance of a broad approach to the relevant issues; that is, an approach that takes into account the stages of research, development, manufacture, marketing and sale of drugs (for brevity's sake, we refer to the collection of these stages as the 'production' of drugs).

The following four facts demonstrate the urgency of creating a substantially different system of drug production and, at the same time, which direction such a change should take. First, there is the unsustainable growth of the public costs of prescription medicines. An important cause of this growth are the monopolistic patents granted to Big Pharma. Second, high drug prices provide a strong incentive for continuing corruption and abuses in the form of misrepresenting the safety and effectiveness of drugs and encouraging their use in situations where they may not be appropriate. Third, the current system is one in which the (mostly big) pharmaceutical industries make excessive profits (much larger than what is usual in other commercial businesses), while they pay hardly any tax on their profits. Fourth, a substantial part of the entire system of drug production is paid by public tax money, through a range of different contributions by national governments and governmental institutions. The latter fact, however, does not have a mitigating effect on the excessive drug prices the public has to pay in their hospitals and pharmacies. The result is that the public pays twice for its medicines: first, via its significant financial contributions to the various stages of the drug production system and, second, for generally overpriced and often excessively expensive medicines.

Our conclusion is that these facts require and justify a shift in our policies for drug production: from privatization through patents to medical research in the public interest. In the first section of the paper we demonstrate that abolishing medical patents is scientifically, socially and morally preferable. The second section argues that it is also economically and financially profitable: it will lead to strongly reduced prices for medicines at pharmacies and hospitals. In the final section we introduce and explain a concrete model of how to do medical research without patents in a way that is socio-politically and organizationally practicable.

In this paper, our primary focus is on medical research in wealthier countries. But of course, the far greater affordability of generic prescription drugs in a system without patents will also be to the advantage of low and middle-income countries. After all, it is the people of these countries who suffer most from the current monopolistic system.

Stricter Standards for Causal Language in Social Science
Nikki Weststeijn

Within biomedical science there is general agreement about the idea that observational studies by themselves do not warrant an immediate causal conclusion. Randomized controlled trials do allow scientists to draw causal conclusions in the papers reporting the trial. In the social sciences, the standards for causal inference are less clear. While there is awareness of the limitations of observational methods, observational studies with causal conclusions are published in leading journals all the time.

In the social sciences, randomized trials are possible, but compared to biomedical science, it is harder to control the environment properly and it is (even) less likely that results generalize to domains outside of the test group. Faith is put into quasi-experimental methods, which include instrumental variable approaches, regression discontinuity designs and difference-in-difference models. While these methods simulate an experimental set-up, they remain observational in nature. They lack randomization and require some assumptions to be made before a causal conclusion can be drawn.

With quasi-experimental methods, in most cases there is no way to know that all the assumptions required for the argument actually hold. Thus, a causal conclusion does not directly follow. We can contrast this with the situation of an ideal RCT, which, as N. Cartwright (2007) has shown, allows a deductive inference to a causal conclusion. As J. Reiss (2007, 2023) has argued for the case of instrumental variable approaches, quasi-experimental methods may vouch for a causal conclusion, but they are no 'clinchers' like RCT's are. While it is unrealistic to require a deductive argument as an answer to all scientific questions, it seems only sensible to be truthful about what we can infer from research results. Thus, I argue that in the case of quasi-experimental methods, scientists should refrain from drawing definite causal conclusions.

The point for this talk is not to enforce the idea that we should only depend on randomized controlled trials in our quest for causality. On the contrary, I agree with part of the message of evidential pluralism (Russo: 2007): that the combined use of different types of evidence (experimental, observational and mechanistic) is likely to increase the trustworthiness and the generalizability of causal conclusions. My point is *to re-emphasize* that, as a scientific community, we should be more careful and specific with causal language then we are now. Just like we cannot simply generalize causal statements inferred from RCT's to the domain outside the test-group, we cannot point-blank draw causal conclusions based on observational methods either. Observational methods should be valued for their own strengths, such as higher external validity.

A more careful and precise use of causal language would hopefully improve the trustworthiness of science and ensure that journalists, policymakers and others interested in reading scientific articles, will not be led astray by premature and unjustified claims.

References
Cartwright, N. (2007). Are RCT's the gold standard? *Biosocieties* **1.1**: 11–20.
Reiss, J. (2007). *Error in Economics: Towards a More Evidence-Based Methodology.* Routledge.
Reiss, Julian (2023). Theory and evidence in economics. In Maria Lasonen-Aarnio & Clayton Littlejohn (eds.), *The Routledge Handbook of the Philosophy of Evidence*. New York, NY: Routledge.
Russo, F. & Williamson, J. (2007). Interpreting Causality in the Health Sciences. *International Studies in the Philosophy of Science*, **21.2**, 157–170.

Mitigating the Theory Crisis in Psychology: Why Formalization needs Theories of Explanation

Luiza Yuan

In this paper I offer a diagnosis of the 'theory crisis' in psychology, a key concern in current methodological research in psychology, and propose a way forward by applying insights from philosophical theories of scientific explanation. The crisis, characterized by a lack of cumulative progress in theory development, stems from two major issues: (**i**) weak inferential links between theoretical constructs and empirical hypotheses, making theories difficult to test (Oberauer & Lewandowsky, 2019), and (**ii**) a focus on describing effects rather than explaining them (Cummins, 2000; Van Rooij & Baggio, 2021).

In recent years, many methodologists propose *formalization* — the translation of verbal theories into mathematical or computational models — as *the* solution for improving the precision and testability of psychological theories and explanations (Borsboom et al., 2021; Fried, 2020; Lange *et al.*, 2024; Oberauer & Lewandowsky, 2019; Robinaugh *et al.*, 2021; Van Rooij & Baggio, 2021).

This 'formalization wave' has, however, faced criticism for failing to address the above- mentioned fundamental issues underlying the crisis, specifically:

a) formalization alone does not resolve high degrees of contrastive and holistic underdetermination in psychology, where compatibility of empirical evidence with many competing theories and large number of auxiliary assumptions handicaps theory-testing (Oude Maatman, 2021);
b) psychology's lack of conceptual clarity and robust phenomena hampers adequately constraining verbal and formal theory development (Eronen & Bringmann, 2021); and
c) vague guidelines for how verbal theories in psychology should be formalized (Lange *et al.,* 2024), combined with the aforementioned challenges (a) and (b), risk producing divergent formalizations of the same verbal theories, resulting in further proliferation of incomparable or untestable claims.

I argue that current formalization approaches pay insufficient attention to the structure, power, and aims of scientific explanation, and that this omission is one of the key issues underlying challenges (a)–(c). I thus propose, and illustrate with scientific examples, that a fruitful way forward for addressing challenges (a)-(c) is by explicitly informing formalization endeavors by insights from philosophical theories of scientific explanation and explanatory power.

Current (discussions on) formalizations proceed directly to a formalization of key posits of a verbal theory, without first specifying what type(s) of explanation can be derived from the theory, for which aims these explanations are procured, and which type of explanation-seeking questions are intended to be answered by them, steps that are vital for constraining theory development. A case in point is an example formalization of the Attitude Similarity Breeds Liking (ASBL) Theory (Lange et al., 2024). I argue that in this formalization, this omission leaves researchers unable to judge whether a formal model improves upon its verbal counterpart or merely repackages existing ambiguities and problems of underdetermination.

By grounding formalization in a theory of explanation, such as the counterfactual theory developed by Hitchcock and Woodward (2003), psychologists would be better able to clarify which dependency relations are explanatory and relevant, to procure testable formal theories, and to assess the comparative strength or power of different formal theories. I illustrate these benefits by relating the discussion to the formalized ASBL model.

References

Borsboom, D., *et al.* (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science*, **16.4**, 756–766.

Cummins, R. (2000). 'How does it work?' vs. 'What are the laws?' Two conceptions of psychological explanation. In: F. Keil and R. Wilson (eds), *Explanation and Cognition* (pp. 117–145). Cambridge, Massachusetts, USA: MIT Press.

Eronen, M. I., & Bringmann, L. F. (2021). The theory crisis in psychology: How to move forward. *Perspectives on Psychological Science*, **16.4**, 779–788.

Fried, E. I. (2020). Theories and models: What they are, what they are for, and what they are about. *Psychological Inquiry*, **31.4**, 336–344.

Hitchcock, C., & Woodward, J. (2003). Explanatory generalizations, part II: Plumbing explanatory depth. *Noûs*, **37.2**, 181–199.

Lange, J. *et al.* (2025, January 10). *A checklist for incentivizing and facilitating good theory building.*

Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review*, **26**, 1596–1618.

Oude Maatman, F. (2021, July 12). Psychology's theory crisis, and why formal modelling cannot solve it. https://doi.org/10.31234/osf.io/puqvs

Robinaugh, D. J., *et al.* (2021). Invisible hands and fine calipers: A call to use formal theory as a toolkit for theory construction. *Perspectives on Psychological Science*, **16.4**, 725–743.

Van Rooij, I., & Baggio, G. (2021). Theory Before the Test: How to Build High-Verisimilitude Explanatory Theories in Psychological Science. *Perspectives on Psychological Science*, **16.4**, 682–697.

Dutch wolf

Jack Nicholson playing a werewholf (1994)

The Dutch village Wierden: a wolf and a pack of local elderly (2024)