The Fourth Conference of the
East European Network
for Philosophy of Science

EENPS 2022

Book of Abstracts

University of Tartu
Tartu, Estonia
August 17-19, 2022

Contents

Keynote talks

Inference to the Best Explanation and Disagreement

Lukáš Bielik
Comenius University Bratislava

Inference to the Best Explanation and Disagreement

Lukáš Bielik

Comenius University Bratislava

Disagreement forms a natural part of our communicative interactions. The paradigmatic case of this phenomenon is the so-called doxastic disagreement. Doxastic disagreement arises when some subject S1 believes that *p* (where '*p*' is a proposition expressing the content of her belief), while another subject S2 believes that $\neg$ *p*, or when S2 *does not* believe that *p*. This kind of disagreement (or its related forms) is the subject of interest in both epistemology and argumentation theory. While epistemology is concerned with the question of what attitude a subject should take, in terms of rationality, if she finds that her peer disagrees with her, argumentation theory examines the different kinds of defeaters that can be represented at the level of arguments.

Disagreement at the level of beliefs is not the only kind of disagreement that has been of interest to philosophy. The disagreement that is the subject of much discussion and disputes in the philosophy of science concerns, among other things, the preferred (recommended) methods and procedures of scientific inquiry, theoretical values, ontological and epistemological assumptions, and the broader methodological background of particular philosophical positions.

In this study, I turn my attention to a particular inference procedure, *inference to the best explanation* (IBE), which characterizes a particular conception of confirmation, and I pay close attention to those parameters of IBE which may be subject to disagreement. For if IBE (in some form) is to function as an inferential rule by which we test, select, or evaluate certain hypotheses, this requires that we agree on certain parameters of IBE (at least in a given context of inference). Expressed equivalently, if we disagree on some parameters of this inferential procedure, then IBE does not allow for an adequate way of testing, selecting, or evaluating the hypotheses under consideration.

I will proceed as follows: in Section 2, I briefly introduce IBE in several variants that dominate current debates. I choose a sufficiently general scheme, which provides the template for the analysis in further sections. In Section 3, I present four generic types of parameters that may be subject to disagreement over IBE. In particular, in Section 3.1 I focus on the kind of situation where two potential subjects S1 and S2 disagree about what constitutes the evidence E that needs to be explained (by a certain hypothesis). Section 3.2 analyzes the disagreement in the case where subjects S1 and S2 are working with different sets of explanatory hypotheses. The disagreement over the space of explanatory hypotheses may - but need not - be independent of another parameter: a model of best explanation. In Section 3.3, I note two basic components of the model of best explanation underlying IBE: (a) the general conception of explanation and (b) the weighting of theoretical virtues that can be used to order multiple explanatory hypotheses into a particular sequence. Finally, in Section 3.4 I discuss situations where subjects S1 and S2 disagree about how to evaluate (or qualify) the resulting hypothesis that stands out in the conclusion of IBE. Section 4 highlights the possibilities of overcoming the disagreement with respect to these parameters. In particular, I show that while disagreement with respect to evidence E is relatively easy to resolve, finding agreement on the other three parameters (the space of hypotheses, the model of best explanation, and the evaluation/qualification of the IBE's conclusion) is much more challenging. On the other hand, the very existence of disagreement over the parameters of IBE may provide a stimulus for further rational discussion between S1 and S2, which could lead at least one of the parties to a broader examination of the underlying assumptions. In Section 5, I revisit the conclusions of the previous parts and note how disagreement on the IBE's parameters is different from disagreement over the parameters of some other inferential rules.

Surrogative reasoning: an artifactual approach

Tarja Tellervo Knuuttila
University of Vienna

Surrogative reasoning: an artifactual approach

Tarja Tellervo Knuuttila

University of Vienna

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

What Social Epistemology Can Learn from Philosophy of Science

Helen E. Longino
Stanford University

What Social Epistemology Can Learn from Philosophy of Science

Helen E. Longino, C.I. Lewis Professor, emerita

Stanford University

This talk distinguishes between minimal and maximal forms of sociality. The talk will argue that a maximal form of sociality characterizes some significant thinking in philosophy of science, focusing on epistemic interactions that are neither reducible to individual actions, nor susceptible to representation in formal networks.

Abstracts

The Unity of Scientific, Technical and Ethical Reason

Ken Archer

The Unity of Scientific, Technical and Ethical Reason

Section: History, Philosophy and Social Studies of Science
Keywords: Phenomenology, History of Science, Philosophy of Science, Technology Ethics, Husserl

Short Abstract

The central argument of this paper is that, underlying the seemingly intractable ethical challenges posed by modern technology - from AI to genetic engineering - is a crisis in the sciences, and that any meaningful progress in the former requires understanding and reckoning with this latter crisis. This crisis of the sciences was articulated by the philosopher Edmund Husserl in the 1930s, but, due to both historical contingencies as well as the academic split between humanities and natural sciences that Husserl denounced and that left Husserl's later work without a proper home, the impact of Husserl's argument on both ethics and science has been muted. This paper seeks to re-articulate this argument and to bring it into dialogue with present day science and technology, focusing in particular on AI ethics.

Long Abstract

The central argument of this paper is that, underlying the seemingly intractable ethical challenges posed by modern technology - from AI to genetic engineering - is a crisis in the sciences, and that any meaningful progress in the former requires understanding and reckoning with this latter crisis. This crisis of the sciences was articulated by the philosopher Edmund Husserl in the 1930s, but, due to both historical contingencies (Husserl was prohibited from publishing as a Jew in Nazi Germany) as well as the academic split between humanities and natural sciences that Husserl denounced and that left Husserl's later work without a proper home, the impact of Husserl's argument on both ethics and science has been muted. This paper seeks to re-articulate this argument and to bring it into dialogue with present day science and technology, focusing in particular on AI ethics.

Contemporary ethics discourse is shaped fundamentally by the modern divorce of the humanities, including ethics, from the sciences, both natural and information sciences, and technology. This divorce casts science as instrumental reasoning, isolated from ethical considerations, and technology as the application of such reasoning. The role of ethics, as a result, is conceived as orienting science and technology from the outside.

Advocates for AI ethics, as an example, tend to teach a set of exogenous ethical principles – fairness, accountability, trust, privacy - that statisticians and engineers are expected to apply. The implication of such advocacy is that, whereas AI systems may not be morally neutral, those who build them are applying instrumental, calculative reasoning to build towards a design and must be taught the ethical ramifications of their finished designs. This formalist framing of AI unwittingly adopts a self-understanding of statistics work that AI ethics advocacy should in fact resist. By buying into formalist assumptions around probability, AI ethics perpetuates the self-understanding of technical work within AI as an applied science which allows little room for agency, such that builders of AI systems have agency primarily in the decision whether to build something.

The divorce of ethics and science, a divorce mutually agreed upon by advocates and skeptics of modern technology, thus perpetuates the ethical crises that it seeks to solve, leading to the intractability we experience today. However, this divorce is not actually true to the history or practice of the technologies and sciences, including AI and statistics, posing these ethical challenges. As Husserl demonstrates, both the humanities and the sciences are grounded in a life-world of intersubjective experience from which the ethical dimension of science and technology emerges.

Husserl develops this understanding of the life-world primarily through a phenomenological analysis that discloses the intersubjective dimension of lived experience, which includes technical practices. Husserl is deeply concerned with demonstrating how the subjective core of lived experience - the mineness of the stream of experience - opens us up to a field of others with shared experiences, in order to constitute the intersubjective objectivity of this experience. In this way, lived experience is characterized by a co-constitution of self, of others, and of objectivity, that cannot be reduced to any one of these three constitutions.

The sense of a single world that is common to all, of an identity and sameness of objects across present and past experiences, experiences of oneself and others, is the basis for the intersubjective constitution of a shared world of possible objects on the basis of which alterity - discrepancies in individual experience - are resolved. The particular mode of intersubjective constitution of a shared world, the basis for resolution of alterity between individual lived experiences, is the critical ethical dimension of science and technology identified by Husserl, overlooked by present-day ethics, and developed in this paper.

The ongoing resolution of alterity through reconstitutions of a shared world gives a teleological orientation to the world in the direction of greater abstraction. The most transformational of such abstractions is that of number, which emerges through the technical practices of measuring, of determining more and less, and which suggests the idealization of an infinite quantity through which the world becomes more mathematical and more scientific.

Husserl distinguishes between an authentic mathematization which is grounded in the meaning structure from which it emerges, and a sedimented mathematization that replaces the life-world of meaning with an idealized reflection, through which experience is now constituted. This substruction of the lifeworld with an idealized schema creates a crisis in the sciences - which are now alienated from their rational basis in practical experience - and a crisis in the practices of daily life, including technical practices - which are now reduced to instrumental techniques of idealized science.

This phenomenological account of the meaning formation and sedimentation of the sciences leads natural to an intentional history of the sciences, one which Husserl provides in broad brushstrokes, with an account of the authentic discovery of number and ideal forms in ancient Greece and their subsequent sedimentation in Galilean science leading to the present-day crisis of the sciences.

However, this intentional history of the formation of intersubjective meaning of the sciences from their basis in technical practices is left largely undeveloped. This paper thus seeks to desediment modern technology, focusing primarily on AI and statistics and their intentional history.

Statistics originated out of classical probability, which was invented in the late 16th-early 17th centuries in Western Europe. The development of classical probability was not a development of pure mathematics. The mathematical methods used in classical probability were already well-known, and it wasn't until the mid-20th century that statistics was formalized in terms of pure mathematical axioms.

So, if statistics was not a mathematical development, what motivated the development of classical probability, without which we would have no statistics, no clinical drug trials, no eugenics, no AI?

The critical backdrop to the 17th century emergence of classical probability was the conviction that civic and commercial order comes not from a transcendent order reflected at different levels of being and enforced by the Church, but from the mutual, intersubjective recognition of men as reasonable through contracts. The specific practices for which classical probability was developed were initially games of chance, followed by other types of aleatory contracts such as insurance, annuities and returns on investments.

The justification of this conviction required an account of how reasonable men engage in cooperative practices in the face of chance and uncertainty, an account that doesn't appeal to fortune as part of the order of things. The ambition of the early mathematicians of probability was to uncover and describe in formal terms the unconscious intuitions of reasonable men. Such a recognition of the universal reasonableness of men, it was believed, would ensure a new, secure basis for social order free of the conflict and skepticism that defined the 17th century.

This paper develops the intentional history of statistics through its sedimentation and crisis, which led to the ethical crises of eugenics, the replicability crisis of social science and the harms of AI, while desedimenting statistics as a technical practice designed to facilitate rational cooperation, the ethical height of which can be seen in clinical drug trials and public health.

This intentional history serves as a model for other intentional histories of modern technology which can similarly desediment their meaning formation and reground their sciences in the intersubjective lifeworld from which they initially emerged.

The Reflective Equilibrium of Intended Models

Nicola Bonatti

Section: Formal Philosophy of Science and Philosophy of Mathematics

Title: The Reflective Equilibrium of Intended Models

Abstract

Most of the literature agrees that the categoricity of Peano arithmetic, Tarski analysis and the quasi-categoricity of Zermelo-Fraenkel set theory fail to demonstrate that there is a unique structure corresponding to our informal mathematical practice. It is argued that the indented models of such theories are not determined by the categoricity theorem alone, rather by the assumption of a special class of axioms, called *'extremal axioms'* – such as the axioms of Induction, Continuity and either Constructibility or Large Cardinals. The leading idea is that extremal axioms imply a *reflective equilibrium* between the informal beliefs concerning the subject matter of a theory and the formal resources adopted to formalize it. Moreover, the dynamic element of the reflective equilibrium highlights that the construction of non-standard models is adopted to either support or revise our informal beliefs concerning, respectively, natural numbers and real numbers.

Keywords: intended model, extremal axioms, reflective equilibrium, non-standard models.

Introduction

Mathematical theories might be distinguished into those that are about a specific subject matter (called *'foundational'* theories) – such as Peano arithmetic, Tarski analysis and, possibly, Zermelo-Fraenkel set theory – and those that are instead designed for different mathematical branches (called *'algebraic'* theories) – such as group, ring and graph theory. Both foundational and algebraic theories support a structuralist view of mathematics. For instance, the intended structures of foundational theories are usually understood in terms of isomorphism-types, in the model theorist's sense. However, while algebraic theories do not rise particular epistemological challenges, foundational theories have to face the *reliability* challenge first formulated by Field (1989): how do we know that foundational theories are about specific structures (i.e. isomorphism-types)? In this talk, I will claim that knowledge of intended structures is obtained through the *reflective equilibrium* between the informal beliefs concerning the subject matter of a theory and the formal resources adopted to formalize it.

Reflective equilibrium (RE) is a coherentist account of epistemic justification tracing back to the work of Goodman (1952), according to which knowledge of some specific domain is reached through a process of mutual adjustment among particular judgements and general principles. In this talk, I will argue that the RE of intended models is obtained through the assumption of a particular class of axioms, called *'extremal axioms'*, which comprehends the axiom of Induction in Peano arithmetic, the axiom of Continuity in Tarski analysis and either the axiom of Constructibility or Large Cardinals in Zermelo-Fraenkel set theory. Extremal axioms specify a condition of maximality/minimality on the class of models $\mathfrak{C}$ defined by the axiomatic theory $\mathcal{T}$ which – given some additional constraints – implies the categoricity of $\mathcal{T}$, see Carnap et al. (1936). For instance, as the axiom of Induction implies that the natural numbers are the *minimal* set closed under the successor function, so the axiom of Continuity implies that the real numbers are the *maximal* Archimedean ordered field.

More precisely, let 'ϕ' be the either First or Second-order extremal axiom of Induction/Continuity/etc. and let '$\mathcal{T}_\phi$' be the axiomatic theory comprehending ϕ, namely Peano arithmetic/Tarski analysis/etc. Then, the coherence of our judgements about the intended models of foundational theories – such as natural numbers, real numbers and, possibly, the set theoretic universe – is obtained trough the RE between the belief that *'It is true that ϕ'* and the belief that *'There exists a class of models $\mathfrak{C}$ satisfying $\mathcal{T}_\phi$'*. Clearly, the two beliefs correspond to, respectively, the judgment and principle of the RE method – more on this below. The RE between the judgement and principle is reached once $\mathcal{T}_\phi$ turns out to be categorical. For example, someone might start by judging the axiom of Induction as a 'self-evident truth'. Having considered the existence of non-standard models satisfying the First-order theory of Peano arithmetic, she revises the formalization of the Induction axiom (and, thus, of Peano arithmetic) so as to achieve a categorical theory. Therefore, the assumption of extremal axioms is essential to obtain the RE determining the intended models. I will support the proposed framework arguing that the assertion of ϕ and the declaration of the class of models $\mathfrak{C}$ satisfying $\mathcal{T}_\phi$ meet the epistemic *desiderata* for judgements and principles. Indeed, while judgements should posses an epistemic standing of initial credibility, principles should achieve some epistemic goals which motivate the transition from judgements in the first place – see Daniels (2020). Finally, I will consider a possible objection concerning the background theory of Second-order logic and its full semantics adopted to establish the categoricity result.

In the last part of my talk, I will apply the proposed framework to the case studies of arithmetic and analysis. I will claim that the mutual adjustments between judgement and principle highlight the epistemic contributions of the construction of non-standard models. More precisely, in the case of arithmetic, the existence of non-standard models prompted for the Second-order formulation of Peano arithmetic, thus supporting the initial plausibility of the extremal axiom of Induction. Instead, in the case of analysis, the existence of non-standard models prompted for the replacement of the axiom of Continuity with another extremal axiom of maximality (i.e. Veronese principle), which applies to both Archimedean and non-Archimedean models. The revised theory is categorical, thus describing the unique model up to isomorphism of the surreal numbers. In this sense, the construction of non-standard models revised our informal beliefs about analysis, concluding that the arithmetic continuum (i.e. the real numbers) belongs to the absolute continuum (i.e. the surreal numbers).

References

Carnap, R., Bachmann, F., and Bohnert, H. G. (1936). On extremal axioms. *History and Philosophy of Logic*, 2(1-2):67–85.

Daniels, N. (2020). Reflective Equilibrium. In Zalta, E. N., editor, *The Stanford Ency-*

clopedia of Philosophy. Metaphysics Research Lab, Stanford University, Summer 2020 edition.

Field, H. H. (1989). *Realism, mathematics, and modality.* Blackwell Oxford.

Goodman, N. (1952). Sense and certainty. *The Philosophical Review*, pages 160–167.

Quietism towards Newman’s Objection to Structural Realism

Kosmas Brousalis

Quietism towards Newman’s Objection to Structural Realism

a. General Philosophy of Science

Abstract

Along with the plethora of the admittedly unsuccessful (purely) structuralist attempts to overcome ‘Newman’s objection’ (NO) to Epistemic Structural Realism (ESR), there is a parallel trend in the literature. That is, the adoption of what I call a *Quietist Stance*, which consists in the suggestion that, although NO is inescapable, it doesn’t have the significance that it is usually given. After examining various relevant proposals, I advance a Quietist Stance based on the classic ‘Restricted Quantifier Response’ (to NO) and ultimately consisting in the following disjunctive assertion: either NO is not a problem *particularly* for ESR or it seems *not* to be a problem for a form of epistemologically motivated, ‘slightly impure’ ESR.

Keywords: Scientific Realism, Structural Realism, Newman’s Objection, Natural Kinds, Quietism

Extended Abstract

A core thesis shared by all forms of Epistemic Structural Realism (ESR) is that we cannot have “non-extensional knowledge of the unobservables” (Frigg & Votsis, 2011: 234). This thesis has become the target of the infamous ‘Newman’s objection’ (NO), according to which the mere assertion that the unobservables instantiate a given extensionally characterized structure is trivially true (granted a cardinality constraint) (Newman, 1928; Demopoulos & Friedman, 1985; Psillos, 1999; Ketland, 2004; Ainsworth, 2009). It is claimed that the only possible way for ESR to avoid the accusation of triviality is to renounce its core tenet, namely the purely extensional characterization of the unobservables.

This talk stems from the observation that along with the plethora of the admittedly unsuccessful (purely) structuralist attempts to show that NO is not sound, there is a parallel trend in the relevant literature; namely, the adoption of what I call a *Quietist Stance* towards NO. Philosophers adopting this stance claim that, although NO is indeed inescapable, it does not have the significance that it is usually given.

Quietism towards NO comes in two flavors. According to what I call the *Robust Quietist Stance*, “Newman’s objection is no trivialization of [ESR], but a compact description of its very point” (Lutz, 2017: 2)—it “simply highlights a consequence of [ESR] that is essential to it” (Worrall, 2007: 152). In respect to this suggestion, I’ll argue that it seemingly rests on a misconception; viz. that the crux of NO is the fact that the domain and the extensions of the relations constituting a supposed concrete structure *S* are underdetermined by the abstract structure whose isomorphism class *S* falls into.

According now to the *Moderate Quietist Stance*, NO is fatal *only* for forms of *pure* ESR. But, if we take into account that “in reality there has never been such a purely structuralist view of theories” (French & Saatsi, 2006: 557), NO loses its bite. What saves ESR from triviality is the imposition of some “additional, intensional constraints” (Bueno & Meier, 2019: 53) on the theoretical content captured by a theory’s Ramsey-sentence (RT), which go “well beyond the mere formal, logical structure of the unobservable world” (French & Saatsi, ibid. 551). Therefore, although “pure structuralism has to be given up” (Bueno & Meier, ibid. 53), a form of *impure* ESR is still on the table.

Apparently, the crucial question arising here is: *how much impurity can ESR afford?* In the course of this talk I examine the prospects of such an impure ESR by looking for a proper balance between the 'minimum intensional content' needed for escaping triviality and the motivations for adopting ESR in the first place.

After reviewing some specific suggestions of the Moderate Quietists, I'll argue that, albeit on the right track, they eventually impose 'too much intensional content'— to such an extent that ESR vanishes. I suggest that the most appropriate move in the service of Moderate Quietists is a classic one: the so-called "Restricted Quantifier Response (to NO)" (RQR), which (quite ironically) has been proposed by the major critic of ESR, namely S. Psillos (ibid.). RQR can be articulated by introducing the interpreted 2nd order predicate 'x is a natural property' (Nx) into the familiar RT, hence taking: $\exists x_1 \ldots \exists x_\nu(\mathrm{T}\,(O_1 \ldots O_\mu, x_1 \ldots x_\nu) \wedge \mathrm{N}x_1 \wedge \ldots \mathrm{N}x_\nu)$. This 'impure RT' is not trivially satisfied, for it is not a priori guaranteed that the structure generated is isomorphic to the *natural structure* of the relevant part of the world.

Nevertheless, Psillos rejects this 'slightly impure' ESR, on the basis of a 'slippery slope' argument of the form 'since you accepted the intensionally interpreted predicate Nx you will have to accept more'. He claims that in order to assert that a relation is natural "we have to go beyond structure and talk about *what* these relations are, […] [for] if one begins with the structure, then one is in no position to tell *which* of the relations one studies and *whether* or not they are natural" (ibid. 66, original emphasis).

However, ESRists are not forced to slip on this slope. I'll argue that, in order to judge that some 1st order relations are natural, we don't have to admit an *epistemically prior* intensional grasp. We merely have to consider the role of the corresponding predicates in the derivation of novel empirical predictions. Namely, if those predicates were corresponding to relations that are the result of gerrymandering, then RT's capacity to bring about novel predictions would be a 'miracle'. Differently put, the 'No-Miracles Argument' (NMA) suffices to do the job for (impure) ESRists, provided that they (*qua realists*) accept its soundness; ESRists don't have to rely on the content that they (*qua structuralists*) wish to renounce.

A similar suggestion has been recently advanced by R. Smithson (2017) who restricts the range of the 2nd order variables via the invocation of the predicate 'the fact that z is such that the NMA provides evidence for it' (§4.1). I think, though, that Smithson's account faces some serious difficulties, which I aim to reveal.

In closing this talk it should be acknowledged that the sketch of impure ESR emerging is possibly subject to several objections. A particular one is fundamental, hence it has intentionally been left occupying the last part of the talk: *why* is it any more problematic to assert that, say, 'x is a proton' than 'x is a natural property'?

In dealing with this question, everything hinges upon one's motivations for adopting ESR. If these are semantic, then ESRists will seemingly (though not inevitably) have a hard time (non ad hocly) justifying the required semantic discrimination and they'll probably have to surrender to NO. However, I'll note that, under the semantic concerns in play, NO is ultimately converting into a parallax of Putnam's Model-Theoretic Argument (Hodesdon, 2013), which is a problem *not specific* to ESR, but one that *all* realists must face. If, on the other hand, the motivations for adopting ESR are epistemic (and granted that we have *somehow* surpassed the semantic issues) then ESRists' task is arguing that we have to withhold belief in the surplus content expressed by the intensionally interpreted T-terms: 'believe solely that the latter pick out *natural-but-don't-know-what* relations'. This task is prima facie feasible.

Under these considerations, my talk can plausibly be construed as proposing a Quietist Stance based on the following disjunctive assertion: either Newman's objection is not a problem *particularly* for ESR or it seems *not* to be a problem for a form of epistemologically motivated, 'slightly impure' ESR.

References

Ainsworth, P. (2009). Newman's Objection. *Brit. J. Phil. Sci. 60*, 135–171

Bueno, O. & Meier, T. (2019). Structuralism, Empiricism, and Newman's Objection. *Principia, Vol. 23 (1)*, 53-67

Demopoulos, W. & Friedman, M. (1985). Bertrand Russell's The Analysis of Matter: Its Historical Context and Contemporary Interest. *Philosophy of Science, Vol. 52, No. 4*, 621-639

French, S. & Saatsi, J. (2006). Realism about Structure: The Semantic View and Non-linguistic Representations. *Philosophy of Science 73 (5)*: 548-559

Frigg, R. & Votsis, I. (2011) Everything You Always Wanted to Know About Structural Realism But Were Afraid to Ask. *European Journal for Philosophy of Science, 1 (2),* 227-276.

Hodesdon, K. (2013). Structuralism and Semantic Glue. https://www.academia.edu/5442689/Structuralism_and_Semantic_Glue

Ketland, J. (2004). Empirical adequacy and ramsification. *Brit. J. Phil. Sci. 55*, 287-300

Lutz, S. (2017). Newman's Objection is Dead; Long Live Newman's Objection! (Preprint) http://philsci-archive.pitt.edu/13018/1/lutz-newmans_objection.pdf

Newman, M. H. (1928). MR. Russell's 'Causal Theory of Perception'. *Mind 37*, 137–148

Psillos, S. (1999). *Scientific Realism: how science tracks truth.* London, Routledge

Smithson, R. (2017). Newman's Objection and the No Miracles Argument. *Erkenntnis 82*, 993-1014

Worrall, J. (2007). Miracles and Models: Why reports of the death of Structural Realism may be exaggerated. *Royal Institute of Philosophy Supplements, Vol. 61*, 125-154

Equivalence without Indispensability?

Jon Charry

Equivalence without Indispensability?

[author's name here]

[author's institution here]

Abstract. Molinini [5] has argued, *pace* Colyvan [2, 3], that the metric tensor—a paradigmatic mathematical object—is dispensable from canonical explanations of relativistic phenomena. His reasoning points to a motif that has received little attention in the literature on hard-road nominalist strategies and which I will underscore: the theoretical *in*equivalence between a platonist theory T and a dispensing theory T^* is a necessary condition for T^*'s genuinely dispensing with some piece of ontological furniture. Furthermore, I argue that this condition is difficult to satisfy simultaneously with the empirical equivalence of T and T^*.

Keywords: nominalism, indispensability, theoretical equivalence, definability theory

1 Extended abstract

It is well known that Quine's "reluctant platonism" was the result of a commitment to what is now called an *indispensability argument* (IA). Colyvan [2,3] has authored the following standard format which most if not all IAs follow. (1) We ought to be ontologically committed to all and only those entities indispensable to our best scientific theories; (2) Mathematical entities are indispensable to our best scientific theories; (3) Therefore, we ought to be ontologically committed to mathematical entities. Colyvan has also offered the standard account of *dispensibility* operative in most if not all IAs. It states that some Xs are dispensable from theory T if and only if there exists a (suitably attractive) theory T^*, the *dispensing theory*, such that

(a) T^* makes no mention of Xs and
(b) T^* is empirically equivalent to the original platonist theory T.

This definition of dispensability has created a cottage industry in the metaphysical and logical methodology of science and philosophy of mathematics. Platonists, like Quine, who stand behind IAs or Colyvan's notion of dispensability are dialectically passive; their commitment to mathematical objects is defeasible, since it is the result of a reluctant acknowledgement that (suitably attractive) theories which satisfy (a) and (b) have, by their lights, not been forthcoming. Despite this, considerable attention has been paid to such alternatives (the most famous example being Hartry Field's project in *Science Without Numbers* [4]).

The example of a "nominalistic" physical theory which will preoccupy us, and which has only recently entered the philosophical literature, is the *Hungarian project* initiated by Andréka, Németi and their students, summarized in [1]. The Hungarian project culminated in the development of SpecRel, an axiomatization of special relativity theory. Molinini [5] has argued, *pace* Colyvan [2], that SpecRel provides genuinely mathematical explanations of special relativistic phenomena, yet is able to do so without invoking the *metric tensor*, a paradigmatic mathematical entity employed in more traditional, platonist presentations of special relativity theory. Molinini concludes that Colyvan is mistaken about the indispensable role of the metric tensor in explanations of special relativistic phenomena.

Molinini's dialectical move calls attention to an important motif which has gone without comment in discussions of hard-road nominalist strategies: nominalists do not take the *in*-equivalence of T^* and T as a necessary condition for T^*'s having genuinely dispensed with abstracta. In fact, Molinini goes as far as to claim that the formal results of Andréka and her collaborators imply that SpecRel is (fully) equivalent to platonist special relativity. This is a serious mistake. We should not say that T^* has genuinely dispensed with abstracta if it is able to "define" or "recover" (in a precise sense) these entities. I show that this points to a way in which Colyvan's conception of dispensability must be sharpened if it is to capture any meaningful sense of dispensability of theoretical terms.

I argue first that clause (b) in Colyvan's notion of dispensability must be replaced by,

(b*) T^* is empirically equivalent to the original platonist theory T *and* T^* is *not* (fully) equivalent to T.

This added restriction can be motivated by making the notion of "full theoretical equivalence" precise with a few candidate precisifications that have been offered in the literature in recent years. In each case, there is a strong sense in which full, theoretical equivalence between two theories implies that each theory can explicitly or implicitly define or recover structures or pieces of ontological furniture that have been allegedly dispensed with by the other. In this case, there is no useful notion of dispensability available to the nominalist, since nothing has been genuinely dispensed with.

I argue, furthermore, that satisfying (b*) is, in many cases, impossible. In particular, this is impossible in cases that philosophers have cited, including SpecRel vs. traditional special relativity. In such cases, empirical equivalence implies full, theoretical equivalence. Here, I will draw on themes from John Norton's [6] work on underdetermination of scientific theories by evidence. Two upshots include both first a critical reassessment of what it means to dispense with mathematical objects, and second, a reassessment of the role of empirical equivalence in Colyvan's definition.

References

1. Madarász J. Andréka H. and Németi I. Logic of space-time and relativity theory. In Pratt-Hartmann I. Aiello M. and van Benthem J., editors, *Handbook of Spatial Logics*, pages 607–711. Springer, Dordrecht, 2007.
2. M. Colyvan. Mathematics and aesthetic considerations in science. *Mind*, 111:69–74, 2002.
3. Mark Colyvan. *The Indispensability of Mathematics*. Oxford University Press, New York, 2001.
4. Hartry Field. *Science Without Numbers*. Princeton University Press, Princeton, 1980.
5. D. Molinini. Evidence, explanation, and enhanced indispensability. *Synthese*, 193:403–422, 2016.
6. John Norton. Must evidence underdetermine theory? In D.H.M. Carrier and J. Kournay, editors, *The Challenge of the Social and the Pressure of Practice: Science and Values Revisited*, pages 17–44. University of Pittsburgh Press, Pittsburgh, 2008.

Scientific credit and the Matthew effect in neuroscience

Matteo Colombo, Michal Klincewicz and Bente Sinke

Scientific credit and the Matthew effect in neuroscience

Abstract According to the Matthew effect, scientists who have previously been rewarded are more likely to be rewarded again. Although widely discussed, it remains contentious what explains this effect and whether it is unfair. Using data about neuroscientists, we examine three factors relevant to clarifying these issues, namely: scientists' fecundity in supervision, H-index, and the location where their PhD was awarded. We find a correlation between location and H-index, but no association between fecundity and H-index. This suggests the Matthew effect entrenches status hierarchies in the scientific credit system not because of exploitative supervisors but partly because of lucky geographical factors.

Keywords: Matthew effect; Robert Merton; Academic genealogies; H-index; Fecundity in supervision; Geographical factors; Reward structure of science.

1. Introduction According to Robert Merton, "[e]minent scientists get disproportionately great credit for their contributions to science while relatively unknown scientists tend to get disproportionately little credit for comparable contributions" (Merton 1973 [1968], 443). Calling this phenomenon the *Matthew effect* in science, Merton motivated its existence based on interviews conducted by Harriet Zuckerman (1977) with Nobel prize winners in the USA (Merton 1973, 440-5; more recent studies indicate the existence of the Matthew effect in research funding and citations, see e.g., Bol et al. 2018; Wang 2014). Several of such interviews illustrate that when an eminent scientist and a junior scientist co-author a paper the eminent scientist will typically receive more credit than their junior co-author. In the words of two Nobel laureates cited by Merton (1973):

In co-authored papers, "[y]ou usually notice the name that you're familiar with. Even if it's last, it will be the one that sticks." (ibid., 444)

“At the extreme, [eminent scientists] sometimes refuse to co-author a paper reporting research on which they have collaborated in order not to diminish the recognition accorded to their less well-known associates.” (Ibid., 446)

Considering these and other testimonies, Merton (1973, 447-50) explains the Matthew effect in terms of its function in the system of scientific communication where attention is limited. If one can only pay attention to a subset of all relevant published research in some field, then a scientist’s eminence can be a cue to the importance of their work. For Merton, scientific communities would thus tend to pay more attention to papers written by eminent scientists and give more credit to the more senior researchers compared to their junior collaborators for co-authored papers.

According to this explanation, the Matthew effect will especially be pronounced for scientists who are highly fecund in supervising many PhD students compared to scientists with few or no supervisees. Supervising junior researchers is one of the most lasting contributions a scientist can make, enhancing their fame in the field (e.g., Marsh 2017). PhD supervisors are typically senior researchers who co-author papers with their students. And these students, after they graduate and leave their supervisors, often adopt similar research approaches as their supervisors, building on their previous co-authored work (e.g., Liénard et al. 2018). Over time, one expects these dynamics will boost the amount of credit senior researchers with many supervisees are given for their co-authored work.

Although intuitively plausible, however, we do not know whether this suggestion is true, since we have limited and mixed evidence that fecundity in supervision relates to the distribution of credit in science. Some studies indicate that supervisors’ collaborative mentorship predicts supervisees’ productivity, understood as the number of research papers a supervisee submits to academic journals (Paglis, Green, & Bauer 2006); other studies highlight that in a field like mathematics the number of doctoral students supervised by a scientist is correlated with the supervisor's number of publications (Malmgren, Ottino & Amaral 2010), and that in applied physics and electrical engineering the number of publications a PhD student produces is positively correlated with their supervisor's age and number of citations (Heinisch & Buenstorf 2018). But these studies do not address whether a scientist's fecundity in supervising many PhD students predicts increased scientific credit in terms of the impact of the scientist's publications.

Philosophers of science interested in the reward structure of science have also discussed the Matthew effect. While these discussions have clarified how the Matthew effect might allocate credit fairly, in an epistemically good way (Strevens 2006), and how this effect might make it impossible to determine the epistemic consequences of the social stratification of science (Heesen 2017), it remains unclear, however, how supervisor-supervisee relationships might relate to the Matthew effect, and in particular to its fairness and epistemic status.

Here we describe the results of a study we designed to clarify such relationships. In our study, we relied on NeuroTree (https://neurotree.org), a large online database that documents the lineage of more than 50,000 PhD advisor-PhD student relationships in neuroscience since the 1850s, to examine whether a neuroscientist's degree of fecundity in supervision is associated with the amount of credit bestowed on their work. We operationalized *fecundity* with a discounted measure of the neuroscientist's total number of mentees (cf., Methods; and David & Hayden 2012), and we used the *H-index*, which is a measure capturing both the quantity and citations of a scientist's publications (Hirsch 2005), as an index of *credit*.

We started by testing these two hypotheses:

(H1) The fecundity of a mentor at time t is a predictor variable of the fecundity of the mentor's descendants at t+n.

(H2) The H-index of a mentor at time t is a predictor variable of the fecundity of their descendants at t+n.

Somewhat surprisingly, our analyses did not provide us with evidence in support of H1 or H2. This result does *not* mean that there is no Matthew effect in a supervisor's fecundity, but it is in tension with Merton's (1973 [1968]) emphasis on seniority and co-authorship in his explanation of this phenomenon. To better understand why we did not find support for H1 and H2, we set out to explore a second set of hypotheses (H3-H4):

(H3) The location where a scientist obtained their PhD predicts the scientist's fecundity.

(H4) The location where a scientist obtained their PhD predicts the scientist's H-index.

We wanted to test these two hypotheses because we assumed that fecundity partly depends on the geographical location where a researcher is trained. Specifically, we assumed that supervising many students requires a relatively high amount of logistic and financial resources, which are not distributed evenly across different geographical locations. Furthermore, researchers trained in certain hotspot locations, such as for example Ivy League universities in the USA, may accrue more credit for their publications based on the reputation of such institutions. If we did not find a main effect of a neuroscientist's fecundity on their H-index, it may be because this effect can be detected only locally, in specific hotspot locations.

Our analyses provided us with some weak evidence in support of H4, as they uncovered local patterns where geographical location is associated with H-index. Thus, geography might be a mediator of the Matthew effect in science, which would call into question the fairness of this effect, if the location where a scientist ends up being trained is largely a matter of luck.

More generally, our study demonstrates that quantitative analyses of academic genealogies can helpfully inform ongoing debates in the philosophy of science about the workings, fairness and epistemic consequences of the reward structure of science.

2. Dataset and Methods To evaluate our hypotheses, we combined data from NeuroTree (https://neurotree.org 2020), which is an online academic genealogy of neuroscientists, and Google Scholar (https://scholar.google.com/2020). Entirely crowd-sourced and continuously updated by any internet user, NeuroTree includes information about 750,000 neuroscientists and more than one million supervisor-supervisee relationships in around 30,000 geographical locations. Data about supervisor-supervisee relationships in NeuroTree are structured as a genealogical tree, stretching back to the earliest days of academic neuroscience in the twentieth century and earlier. The resulting tree-like data structure provides a unique opportunity to study quantitatively the development of the field of academic neuroscience.

To calculate a neuroscientist's *fecundity,* we followed David & Hayden (2012) in using a discounted sum of the number of connections to any one researcher over *N* generations of supervisees. Specifically, a scientist's fecundity score was calculated with the formula:

$$\textit{Fecundity Score} = n_1 + \gamma n_2 + ... + \gamma n_m$$

where n_1 represents the students directly supervised by a researcher, n_2 represents the students of n_1, n_m represents the number of students through m successive generations, and γ is a discount factor to avoid a bias in favor of older researchers (David & Hayden 2012). This score was computed up to the 5th generation for each researcher to test H1 and H2.

H-index is included in the NeuroTree dataset, but unfortunately this variable has 97.06% missing data in that dataset. Therefore, we used Google Scholar (https://scholar.google.com/ 2020) as a separate source of data about the H-index of the researchers in NeuroTree. Since no open-access Google Scholar API exists, we created a custom program to automatically extract H-indexes from the Google Scholar website. A total of 1,383 entries were eventually used in this part of analysis.

We also sourced and added a unique academic identifier ORCID (https://https://orcid.org/ 2020) for each scientist in our dataset with the intention of gathering location-specific information about where their PhD was awarded. After inputting the full name of a scientist, the program activates the scholarly module (https://pypi.org/project/scholarly/2020), allowing a connection with Google Scholar. The custom program then imports all data from the specified scientist.[1] From past education entries (from ORCID and hand-labelling), only relevant PhD degrees and corresponding locations were retained. The locations (listed as strings containing specific organization names) were geocoded using the Nominatim API (https://nominatim.org/release-docs/latest/api/Overview/). This application uses strings to search for corresponding geolocations, as well as the longitude and latitude coordinates required for geographical analysis.

To select entries for geographical analysis, 1,374 entries were found that included registrations of histories of the educational backgrounds in ORCID profiles, and out of those only 342 could be geocoded. These were combined with 41,390 entries that did not have an

[1] While importing the data, a custom search log is created to track the searches and record metadata concerning the imports. By using this search log, unique results are filtered and non-unique instances are deleted, counteracting the non-unique nature of the search input variable (full name of an academic).

ORCID profile listed in the original NeuroTree dataset. The names of these individuals were cross-referenced with the ORCID API and every time a registered name and location linked to that name in the NeuroTree dataset matched the profile information returned by the ORCID API, a record was included into our new dataset. From these 41,390 individuals who were cross-referenced by name, 1,157 locations could be geocoded. So, the final dataset used for analysis included 1,499 (342 + 1157) data points, each corresponding to a neuroscientist. In this final dataset new variables such as geolocation, longitude, latitude, and geometry were added accordingly.

After the data was cleaned and adapted for testing our four hypotheses, we followed a shared pipeline for regression analyses. This includes the standardization of each of the input variables, which allows unbalanced variables to perform better when using predictive models (Pedregosa et al. 2011). After standardizing the data, the test set was extracted from the original dataset. K-fold cross-validation with five splits was used for all regression analyses. To add to the comparability, a set random state was used to retrieve repeatable and comparable results.

2.1 Predictive Models SGDRegressor, Decision Forest, and Lasso models were used for probing the relationships between fecundity, H-index, and geographical location. To determine the success of these models in predicting the value of a target variable based on a given predictor, the following performance metrics were used: Mean Squared Error (MSE), Mean Absolute Error (MAE), Explained Variance (EV), and R2 score (R2). To estimate a proper fit of a model to our data, Explained Variance and R2 were used. Due to a lack of research concerning the subject of this paper, no baseline could be formed for model comparison. Therefore, interpretation of the fit of the models was based on the ability to explain more than a mean estimation for each variable.

2.2 Geographical Analysis We performed a Global Moran I analysis as an indicator of general clustering tendencies across geographical space for both H-index and fecundity. This analysis indicates dissimilar, random, or similar values of a given variable clustering together across geographical space. Specifically, the Global Moran I analysis approximates one general global clustering tendency that is most prevalent across all points. High values clustering to high values, or low values clustering to low values are considered to belong to the same clustering tendency since the observation and its surrounding points belong to the same class. To identify regional or local differences in terms of clustering tendencies, taking into account the specific variable values, a Local Moran I test was performed.

3. Results Our regression models uncovered no statistically significant relationship between fecundity and H-index. In particular, we found no evidence in support of H1 (the best performing model was Random Forest with an R2=0.079) and H2 (the best performing model was a Random Forest with R2 0.085 and MAE 20.20).

In an explorative analysis, we focused on the variable "Collaboration" in the NeuroTree dataset. Because this variable is not coded as a mentor-mentee relation in NeuroTree, it does not directly contribute to the fecundity score of an individual researcher. But we found non-zero Lasso coefficients when looking at whether it predicts a researcher's fecundity score. This indicates that being a research collaborator predicts similar fecundity scores better than being a mentor.

We followed up with a geographical analysis, looking into the potential relationship between the location where a neuroscientist received their PhD degree, fecundity and H-index. Based on geographic analysis, there appears to be an overall weak significant geospatial effect for similar H-index values to cluster together. This result indicates clustering effects of H-indexes do *not* originate from random processes, but from underlying geographical processes.

Regarding total fecundity however, we found no significant relation with geographical location.

For geographical analysis, the selected distance band affects which neighboring datapoints are taken into consideration for the statistical test. As such, the ideal distance band is selected by using the distance radius or computed distance weight that yields the first significant effect, hereby representing an optimal fit for the data. As shown in Table 1, the first significant effect for H-index on a world scale (z-score: 1.743, p-value: 0.045) corresponds to an autocorrelation value of 0.023 and is observed at a distance radius of 1.6 degrees, representing a distance radius of 177.6 km around the observed locations respectively.[2] A Global Moran I's autocorrelation value of 0 indicates a random dispersion of values, and an autocorrelation value of 1 would indicate similar values clustering together across geographical space whereas a score of –1 would indicate dissimilar values clustering together.

The autocorrelation value suggests a tendency for similar values of H-index to cluster together across geographical space. But we should note that the observed significant clustering tendency is very weak due to the respective autocorrelation value only slightly deviating from 0. It should also be noted that the theoretical ideal distance band of 1.6 degrees radius, should be contrasted to the scale of the actual physical distances between data points. For instance, since this distance band was computed on a world scale (including datapoints all over the globe), the radius might appear to be a disproportionately large radius of influence within the context of geographical dispersion between cities in Europe. To illustrate, H-index values from PhD graduates from Utrecht University in the Netherlands would be influenced by nearly any other location in the Netherlands. Within the context of geographical dispersion between cities in the United States, however, a distance of influence of 177.6 km between datapoints could be considered more realistic.

[2] One degree difference longitude or latitude amounts to approximately 111 km.

In contrast to H-index, the Global Moran I analysis did not produce any significant results for total fecundity regardless of distance band value. Despite an optimal distance band (z-score: -0.924, p-value: 0.129) corresponding to a radius of 20 degrees or 2220 km and resulting in an autocorrelation value of -0.003, both p-value and z-score failed to reach statistical significance. Since the null hypothesis indicating that these effects originate from random processes rather than geographical processes cannot be rejected, there is no evidence to support that total fecundity is affected by geographical location.

Table 1: Global Moran I analysis results.

	Optimal Distance Radius	Autocorr.	P-value	Z-score
H-index (World Scale):	**1.6**	**0.023**	**0.045**	**1.743**
H-index (United States only):	**1.4**	**0.027**	**0.039**	**1.917**
Total Fecundity (World Scale):	20	-0.003	0.129	-0.924

Since a Global Moran I analysis is an indicator of general clustering tendencies, it does not take into whether these specific clusters contain high or low values. In order to identify regional differences in terms of clustering tendencies and taking into account the specific variable values, a Local Moran I test was performed. Considering how the Global Moran I test for total fecundity did not yield any significant results, the variable was irrelevant for the subsequent Local Moran I analysis and was not included. Instead, the Local Moran I analysis focused solely on exploring differences in regional or local patterns that may exist regarding the relationship between h-index and geographical location.

Between H-index and geographical location, visualization of the Local Moran I analysis in Figure 1 shows tendencies for high values to cluster together to appear most prominently in

Japan, the Eastern regions of the United States, and Western Europe. While the tendency for high values to cluster together appears to be relatively centralized, the tendency for low values to cluster together appears far more dispersed over the globe. Moreover, this tendency for low values to cluster together appears to outweigh the number of observations in which high values tend to cluster together.

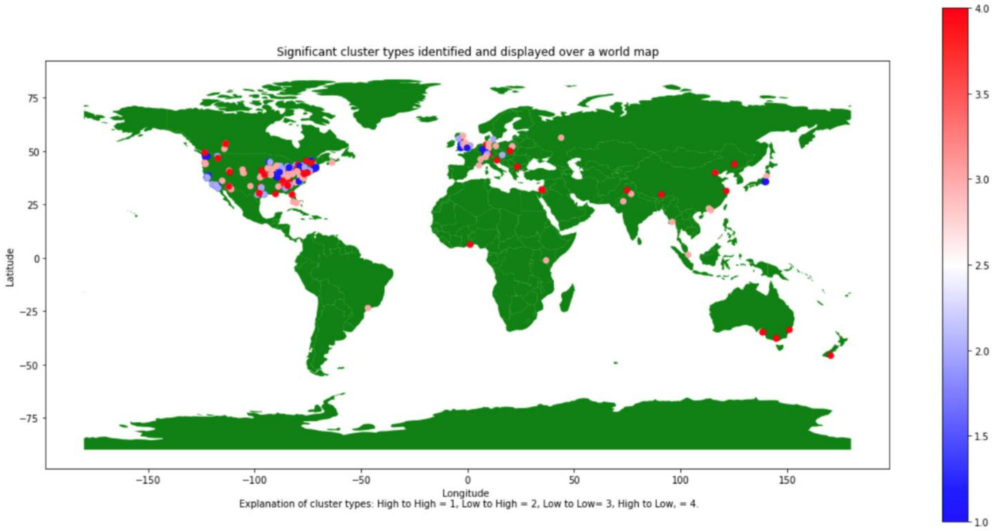


Figure 1: Clustering types displayed on world map. Explanation of cluster types: High to High = 1, Low to High = 2, Low to Low= 3, High to Low, = 4.

Furthermore, the high-to-low clusters marked by the bright red dots appear to be similarly dispersed across the globe. This dispersion contrasts the low-to-high type clusters marked by the light blue dots which appear once again more concentrated in the United States and Western Europe. Contextualizing these clusters, the high-to-low clusters could considerably represent outlier academics with high H-index scores living in remote areas where their general surroundings does not match their exemplary academic credit. The low-to-high clusters could be interpreted as resembling the hierarchal structure of academia in which junior scientists (with low H-index) are drawn to eminent scientists with higher measures of scientific credit, so that they may collaborate with them, trying to advance their academic career. Over time, these junior scientists accumulate credit themselves and attract those with lower H-index scores to engage in collaboration with, starting a new cycle of low-to-high clusters or becoming a high-to-low cluster (depending on the point of reference.)

Regional differences with respect to clustering types are particularly poignant in the United States and Western Europe. While this is biased by sample size, considering how locations in the United States were overrepresented in the geographical analysis dataset, the imbalance also emphasizes how the dataset and to an extend the distribution of credit, appears to have a western (historical) bias.

Overall, these results provide some support to hypothesis H4. The weak observed clustering tendency for either similarly high or similarly low values of H-index to cluster together in space may suggest that credit centralizes geographically in specific places. As such, when predicting the distribution of credit, scientists who are surrounded by others with high H-index values might be more likely to gain increased scientific credit in terms of an increasing H-index in specific locations. But it may also mean that high performing scientists are attracted to these places for other reasons, such as higher salary, better infrastructure, or quality of life.

4. Discussion The main aim of our study was to clarify the nature of the Matthew effect, by examining three factors that might contribute to explain and interpret it, namely: a scientist's fecundity in supervising many PhD students, the scientist's H-index and the location where their PhD was awarded. We tested four hypotheses concerning these factors, based on analyses of a large dataset we compiled with information from NeuroTree, Google Scholar and ORCID.

Our analyses revealed no associations between the fecundity of a neuroscientist at a given time and the fecundity of their mentees at a subsequent time (H1), between a neuroscientist's H-index at a given time and the fecundity of their mentees at a subsequent time

(H2), and between the location where a neuroscientist obtained their PhD and their fecundity (H3). We could only find weak evidence in support of an association between the location where a neuroscientist obtained their PhD and the scientist's H-index (H4). These results have implications both for the understanding of the sociology of neuroscience and of the reward structure of science more broadly.

Firstly, we should be clear that our negative results do *not* show that there are no Matthew effects in neuroscience, especially considering the relatively small sample in our study and poor data quality. In particular, our negative results do *not* support the idea that the reward system in science bestows credit in proportion to a scientist's contribution in mentoring many students, nor do they support the idea that scientists' research impact increases their overall number of mentees.

Secondly, our results contribute a more nuanced interpretation of some existing explanations of the Matthew effect, its fairness and epistemic function. Consider Robert Merton's (1973) explanation of the Matthew effect. Emphasizing the role of seniority in co-authorship, Merton (1973) suggested that the Matthew effect depends on our tendency to pay more attention to more eminent people, which would lead to an unfair allocation of credit to eminent scientists compared to "nobodies" in science. If a scientist's fecundity mediates the possible relationship between differential attention and differential allocation of credit, then we should have found some evidence of a disproportionate higher H-index for researchers with many PhD students, who would presumably co-author with their supervisors and cite their supervisors' work in their subsequent papers; and more generally, we should have found a positive association between a scientist's fecundity and their H-index. But we did not find this evidence; so, the data seems to run in the opposite direction to the testimonies of scientists cited by Merton. Whatever the merits of Merton's explanation, fecundity does not appear to mediate the relationship between differential attention and differential credit to eminent scientists.

Next consider Michael Strevens's (2006) explanation of the Matthew effect, according to which the eminence of a scientist reliably indicates the scientist's trustworthiness; and so, eminent scientist would receive more credit for their research, which will in turn increase their eminence. This explanation relies on the premise that "the reward system in science bestows credit in proportion to a scientist's contribution to society" (164). While Strevens illustrates this premise by alluding to Louis Pasteur's contribution to society in terms of lives his discoveries contributed to save, he does not say much about how we should understand and measure contributions to society.

Our study bears on Strevens's explanation in two ways at least. First, our results undermine the idea of a genius scientist that single-handedly moves the field and single-handedly contributes to society. A model of a reward structure of science that is grounded in this ideal does not reflect the reality of contemporary scientific practice. Science is increasingly done by communities that function differently depending on location; and David and Hayden's (2012) analyses of NeuroTree independently reveal that the field of neuroscience is akin to a cottage industry where small communities of researchers tackle small problems with specialized techniques they master. Second, and more importantly, our study singles out mentorship as one of the most salient contributions scientists from any discipline (including non-applied sciences) can make to society (see, e.g., Malmgren et al. 2010; Marsh 2017; Sternberg 2018). If we spell out Strevens's premise about scientists' contributions to society in terms of mentorship, then our results call that premise into question. The scientific reward system does not seem to take account of the level of fecundity of a scientist in mentoring. If this is true, then our results indicate that Strevens's explanation rests on weak grounds, and his interpretation of the Matthew effect should be reconsidered.

Thirdly, we found that the geographical location where a scientist received their PhD predicts the scientist's H-index. We also found a tendency for either similarly high or similarly low values of H-index to cluster together in geographical space. This suggests that highly promising PhD students who are surrounded by other promising PhD students might be more likely to similarly get recognition in science. But this tendency makes it salient how *luck*, rather than scientific competence or merit, plays an important role in explaining the allocation of credit in science (Heesen 2017). After all, the geographical location of one's PhD is associated not only with reputation, but also with differential availability of financial resources, equipment and infrastructure (see e.g., Cummings & Kiesler 2003; Chariker et al. 2017). Such material differences have less to do with researchers' competence or merit than with historically lucky, social, political, and economic processes, which can in turn disproportionately benefit an institution's reputation. Because of historically path-dependent socio-economic disparities, geographical location might therefore facilitate higher levels of academic success *regardless* of the degree of competence or merit of a mentor or mentee.

While this conclusion is consistent with the testimonies discussed by Merton (1973), which suggest that the Matthew effect violates norms of fair allocation of credit in science, we should not over-interpret it. Although we analyzed a large dataset combining information about academic genealogy and publication impact, we should emphasize its limitations in terms of missing and corrupted data, and a strong bias towards researchers from North America and to a lesser extent Europe. Taking these limitations into account, our overall results indicate that fecundity in supervising many PhD students might not contribute to explain the Matthew effect in neuroscience; geographical location is a more plausible factor. The Matthew effect might thus entrench unfair status hierarchies in the scientific credit system not because of exploitative supervisors but partly because of lucky geographical factors.

References

Bol, T., de Vaan, M., & van de Rijt, A. (2018). The Matthew effect in science funding. *Proceedings of the National Academy of Sciences*, 115(19), 4887-4890.

Chariker, J. H., Zhang, Y., Pani, J. R. & Rouchka, E. C. (2017). Identification of successful mentoring communities using network-based analysis of mentor-mentee relationships across Nobel laureates. *Scientometrics* 111, 1733–1749.

Cummings, J., & Kiesler, S. (2003). Coordination and success in multidisciplinary scientific collaborations. *ICIS 2003 Proceedings*, 25.

David, S. V. & Hayden, B. Y. (2012). Neurotree: A Collaborative, Graphical Database of the Academic Genealogy of Neuroscience. *PLoS ONE* 7(10): e46608. https://doi.org/10.1371/journal.pone.0046608

Heesen, R. (2017). Academic superstars: competent or lucky?. *Synthese*, *194*(11), 4499-4518.

Heinisch, D. P., Buenstorf, G. (2018). The next generation (plus one): an analysis of doctoral students' academic fecundity based on a novel approach to advisor identification. *Scientometrics*, 117, 351–380. https://doi.org/10.1007/s11192-018-2840-5

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences*, *102*(46), 16569-16572.

Liénard, J. F., Achakulvisut, T., Acuna, D. E., & David, S. V. (2018). Intellectual synthesis in mentorship determines success in academic careers. *Nature communications*, *9*(1), 1-13.

Malmgren, R. D., Ottino, J. M., & Amaral, L. A. N. (2010). The role of mentorship in protégé performance. *Nature*, 465, 622–627.

Marsh, E. J. (2017). Family matters: Measuring impact through one's academic descendants. *Perspectives on Psychological Science*, *12*(6), 1130-1132.

Merton, R. K. (1973). The Matthew effect in science. In *The sociology of science* (pp. 439–459). Chicago: Chicago University Press. (First published in *Science*, 159 (1968), 56–63).

Paglis, L. L., Green, S. G., & Bauer, T. N. (2006). Does adviser mentoring add value? A longitudinal study of mentoring and doctoral student outcomes. *Research in Higher Education*, *47*(4), 451-476.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *The Journal of machine Learning research*, 12, 2825-2830.

Sternberg, R. J. (2018). Evaluating merit among scientists. *Journal of Applied Research in Memory and Cognition*, 7(2), 209-216.

Strevens, M. (2006). The role of the Matthew effect in science. *Studies in History and Philosophy of Science Part A*, *37*(2), 159-170.

Wang, J. (2014). Unpacking the Matthew effect in citations. *Journal of Informetrics*, *8*(2), 329-339.

Zuckerman, H. (1977). *Scientific elite: Nobel laureates in the United States*. New York: Free Press.

Getting understanding in multispecies ethnography

Richard David-Rus

Getting understanding in multispecies ethnography

Anthropology is classified in the register of social sciences and humanities though there is split cutting through its body which makes it to be partially claimed by natural sciences. It is not only the case of biological anthropology claimed by such sciences as evolutionary biology, but also other subdomains of anthropological studies which draw in an interdisciplinary way on domains from natural science. Moreover, this seems to take place even in the register of cultural anthropology as recent developments reveal. This fact is not without consequences in what regards the sort of scientific understanding gained in anthropological inquiry – the issue that I will address in this contribution.

I order to do this I will focus on some recent trend in the cultural anthropology known under the name of multispecies ethnography. From a historical point of view understanding in anthropology was claimed esp. in the tradition of understanding as Verstehen, meaning the special sort of understanding in humanities and social sciences. I will argue that the new trend puts a heavy pressure in reconsidering the interpretation of understanding as Verstehen in anthropology. The reliance on natural sciences that is essential in multispecies ethnography such as biology, ethology, ecology or geography diminishes the chances of understanding as Verstehen.

In a first step I will look at the original interpretations of Verstehen taking Martin's reconstruction (Martin 2000) of the Dilthey's classical position. From the three interpretations he distinguishes in Dilthey's work: the reliving interpretation, the reconstruction one and the cultural context, the first two seem problematic from the beginning. The first one was contested as not being a necessary condition even in case of humans since it is not necessary to empathize with the inquired subjects meanwhile the second one is problematic due to required knowledge of inner life that is needed for the reconstruction of the subjects' experiences. The cultural context could be invoked with reference to the animal culture but the concept has a stricter definition than in case human culture as we might identify different specific aspects in different species. Besides we gain understanding on animal cultures by deploying methods from natural sciences as ethnology, ecology and related fields..

In the second step I will analyses directly the way understanding is gained in multispecies ethnography by discussing a recent piece of inquiry in the field - the study of J Hartigan (2021) on wild horses in Galizia, Spain. Hartigan intends to make a consistent contribution to the methodology of multispecies ethnography by arguing "for an ethologically informed ethnography that extends cultural analysis to other social species". I will argue that we might have difficulties in cashing something on the side of Verstehen from his account.

Hartigan studies the annual ritual of shaving the wild horses (rapa das bestas) in Galizia avoiding Geertz's way of taking animals as representations and projections of humans interests and concerns. Rather the author is interested in approaching horses' sociability directly and study the impact of the ritual on it. He applies ethological techniques in his direct observation of his subjects and this might be seen as a first level of gaining understanding. This understanding is not different from any other understanding gained in ethology and so subsumable under a natural science sort of understanding. The specific touch of Verstehen might come on the second level when the author applies concepts from social analysis, from Goffman's theory of social interaction to these data. The author claims that via this analysis horses as social subjects 'engage in ongoing interpretative work in understanding, reproducing and contesting their relationships". Nevertheless we have here rather a metaphorical attribution via a conceptual analogy when claiming the ongoing interpretative work on the part of horses. There are no traces of Verstehen ingredients in this case as they are claimed by more recent theories of Verstehen. Take for example two recent accounts favorable to Verstehen such as Stuber's and Grimm's approaches. The first author (Stuber 2012) is pointing to our capacity of reenactive empathy that helps us grasp the reasons for an action being this way a variant of the reliving interpretation. For Grimm (2016) it is 'understanding-as-taking-to-be-good' that comes from being able ot see or regard the subject's end as good or choiceworthy that makes for the sort of Verstehen kind of understanding. None of these two might be identified in the process of gaining understanding in our case.

As a last step we might take a look from the perspective of the contemporary theories of understanding as advanced in epistemology and philosophy of science. I will take as reference two of them: Wilkenfeld's (2013) and Dellsen's (2020) accounts. According to the first one understanding a phenomena involves representation manipulability meanwhile for the second author it involves grasping a model of phenomenon's dependence relations. Both theories can accommodate our case. In order to make the point for a specificity of the understanding in this case one has to identify the special ingredient of Verstehen in some step of these accounts: either the way one builds representations or manipulates them in social sciences or the models of dependence relations. Though there could be might to identify such ingredients in particular episodes of social research, the case discussed here does not seem to be among such episodes.

References

Dellsén, F. (2020). Beyond Explanation: Understanding as Dependency Modeling. *British Journal for the Philosophy of Science* (4):1261-1286.

Grimm, S. (2016). How Understanding People Differs from Understanding the Natural World. *Philosophical Issues* 26 (1):209-225.

Hartigan Jr, J. (2021). Knowing Animals: Multispecies Ethnography and the Scope of Anthropology. *American Anthropologist*, *123*(4), 846-860.

Martin, M. (2000). Verstehen: The Uses of Understanding in Social Science (1st ed.). Routledge.

Stueber, K. R. (2012). Understanding versus explanation? How to think about the distinction between the human and the natural sciences. *Inquiry*, *55*(1), 17-32.

Wilkenfeld, D. (2013). Understanding as representation manipulability. *Synthese* 190 (6):997-1016.

The Later Wittgenstein, Hinges, and Mathematical Practice

Jordi Fairhurst, José Antonio Pérez Escobar and Deniz Sarikaya

Title: The Later Wittgenstein, Hinges, and Mathematical Practice

Submitted to: f) Formal Philosophy of Science and Philosophy of Mathematics

Keywords: Hinge Epistemology, later Wittgenstein, Philosophy of Mathematical Practice,

Short abstract (<100 words):
In this talk we give a Wittgensteinean account of hinge epistemology for the mathematical practice. Against previous interpretations of Wittgenstein, this account claims that mathematical formulations per se cannot be either epistemic or non-epistemic, but their uses can be epistemic or non-epistemic based on nuances of mathematical practices. We argue that this account of Wittgenstein's hinge epistemology is more faithful to the views of the later Wittgenstein on mathematics. Furthermore, to test the ecological validity of this Wittgensteinian account of mathematical hinges, we contrast it against a biographical account of mathematical work put forward by the Field medalist Terence Tao.

Abstract (<1000 Words Words):
In this talk we give a Wittgensteinean account of hinge epistemology for the mathematical practice. We will be concerned with three interconnected issues: 1) the lack of common properties of hinges across different situations and the implications for hinge epistemology, 2) Wittgenstein's conception of mathematical hinges as rejecting a unified theory of hinges, and 3) an ecological validation of the rejection of such a unified theory of hinges informed by the mathematical practice.

Hinge epistemology is an umbrella-term for a diverse group of epistemological theories about justification and knowledge that expand on Wittgenstein's concept of 'hinges' in On Certainty. This concept can be roughly defined as the fundamental presuppositions of one's worldview which are exempt from exempt from doubt and make it possible for us to perform other epistemic operations (e.g., discovering, justifying, verifying, etcetera). Competing hinge theories have different views on how best to understand hinges, each with different implications for the analysis of our epistemic practices. Non-epistemic theories (see e.g., Moyal-Sharrock 2004, 2016; Stroll 1994; Pritchard 2011) claim that hinges are outside the scope of rational evaluation and lack epistemic properties, i.e., they are neither justified nor unjustified. Meanwhile, epistemic theories (see e.g., Williams 1991; Wright 2004; 2014; Hazlett 2014) claim that hinge commitments are within the scope of rational evaluation and do have epistemic properties (i.e., they can be potentially justified or unjustified), albeit in a non-paradigmatic way.

The standard approach to constructing hinge theories is marked by the assumption that the nature of hinges is uniform and that, consequently, it can only be adequately explained by one global theory. We contend that Wittgenstein's observations on the complexity of, and lack of essential characteristics shared by, hinges together with the variety of their characteristics in different epistemic practices, give us good reasons to break with this assumption. Hinges are complex phenomena whose characteristics do not display a theoretical unity. Accordingly, resorting to Wittgenstein's methodological pronouncements, we develop a piecemeal hinge epistemology where we approach hinges, not by putting forward global theories that claim to explain everything essential about all hinges. Rather, we seek to clarify specific hinges on a case-by-case basis by means of a variety of complementary local models. On this basis, we can rectify existing hinge theories to local models by restricting them to the specific hinges with the characteristics that the theories can clarify.

To achieve this, we set out to investigate whether mathematical hinges OC can be interpreted to be either epistemic or non-epistemic. Importantly, we find parallelisms between the conception of

mathematics in Wittgenstein's late work (LFM, RFM) and that of OC: for instance, both adhere to the Wittgensteinian hallmarks of "meaning as use" and "mathematics as rules of description/petrified empirical regularities". Indeed, others have resorted to Wittgenstein's views on mathematics in order to support the interpretation that, for Wittgenstein, hinges are non-epistemic (McGinn, 1989; Moyal-Sharrock, 2005) but also the interpretation that hinges are epistemic (Kusch, 2016).

By building on the later Wittgenstein's philosophy of mathematics, we challenge both of the above interpretations. For instance, Moyal-Sharrock's account can be challenged on the ground that she mostly refers to the early Wittgenstein's views, the continuity of which with OC is unlikely. As for Kusch's interpretation, while he is aware that Wittgenstein's view of mathematics as "petrified regularities" has a normative grammatical aspect, he seems to suggest that because there is an empirical aspect as well (and hence mathematical hinges can be true or false), then mathematical hinges are epistemic. Yet, as we note, Kusch resorts to a violation of the other Wittgensteinian hallmark, meaning as use, in order to make his claim. He claims that a given mathematical hinge (for instance, formulations like 2 + 2 = 4) may have both epistemic and non-epistemic uses, but for Wittgenstein, it is not symbols (number or letters) that comprise the essence of mathematics, but mathematical uses. Hence, it is not that a mathematical formulation is normative but is also epistemic, and hence mathematical hinges are epistemic, but instead, different uses of a symbol array constitute different mathematics and mathematical meaning (and hence, different mathematical hinges). The formulation does not comprise self-sufficient mathematics, but the mathematical practice "fills the gaps". Different practices employ different kinds of hinges, even if these may share the same symbol array. LFM and RFM support this view extensively, and so does OC (98; 139). Therefore, there is no reason to believe that Wittgenstein conceived mathematical hinges as either epistemic or non-epistemic.

Finally, we explore features of the mathematical practice in order to ecologically validate our interpretation of Wittgenstein's view of hinges. We elaborate on a model by Terence Tao, which says that there are three phases within the biography of a mathematician: a pre-formal phase, a formal phase and a post formal phase. In the first phase we familiarize ourselves with the concepts involved in a given mathematical practice. We then transgress to the other two phases which are constantly alternating. The second phase is a kind of work very close to the axioms/ definition and rules. We argue that this phase is closely related to a non-epistemic use of mathematics. In Phase III, a person resumes informal actions after gaining experience and knowledge of the necessary mathematical context, as well as the ability to formally operate within it. In this phase mathematical formulas are not interpreted formally, but mathematicians employ background theories that yield epistemic uses of the formulas. In a way that makes the phase related to an epistemic use of mathematics. Even more, when our informal arguments eventually become problematic, the mathematician returns to the second phase, discarding every informal 'shortcut' etc. and restricts themselves to the formal corset. This change from II to III phase and vice versa repeats often, which illustrates how mathematics can transition from non-epistemic to epistemic, and from epistemic to non-epistemic. This model exemplifies our Wittgensteinian take on mathematical hinges and is in itself a worthwhile contribution to the study of mathematical practice.

Renovating the Child as Scientist Hypothesis

Mark Fedyk

Title: Renovating the Child as Scientist Hypothesis

Keywords: developmental psychology, philosophy of cognitive science, philosophy of science, abduction, induction, rational constructivism

Short Abstract:
Both children and scientists can solve the problem of induction. This has inspired a literature exploring similarities between the idealised accounts of the cognition of scientists and theories in developmental psychology about how the developing minds of young children operate, where these explorations are typically organised as assessment of the "child as scientist" hypothesis (CATH). This presentation assesses CATH both in light of research in developmental psychology undertaken over the last two decades, as well as work in the philosophy of science that deepens our understanding of the situatedness and normativity of scientific inquiry. On the basis of these two lines of evidence, I propose a substantial renovation of CATH.

Long Abstract:
Both children and scientists are capable of solving the problem of induction. This fact has inspired a literature exploring similarities between the idealised accounts of the cognition of scientists and theories in developmental psychology about how the developing minds of young children operate, where these explorations are typically organised as assessment of this "child as scientist" hypothesis (CATH). The canonical expression of this work is now a quarter century behind us, being Alison Gopnik and Andrew Melztoff's "Words, Thoughts, and Theories" (Gopnik & Meltzoff, 1998). As twenty five years is likely the minimum amount of time necessary for a profound philosophical idea to mature, with this presentation, I propose to offer a philosophical assessment of the Child as Scientist Hypothesis (CATH).

In its original version, CATH depicts young children as learning by successively correcting conceptual deficits driven by a process that looks strikingly similar — especially in its consistency with methodological axioms governing how to update belief using statistical and causal evidence — to rational theory change in science. In developmental psychology, these ideas have helped to inspire roughly two generations of research that have greatly added to our understanding of how children learn. Partly because of this research, my thesis is that CATH needs renovation: we have learned more about children and more about science in the intervening years, and these lessons suggest that CATH at least needs a new coat of paint.

But there have been further developments in the philosophy and sociology of science which portend bigger changes to CATH, work on the foundations and framing of CATH. These developments are accounts of the integrative role that normativity plays in scientific progress, and accounts of scientific practice (and subsequently certain kinds of progress) that, in their explanations of both practice and progress, make in eliminable reference to the situatedness of scientific research. These developments have important logical consequences for what sorts of ideals fit normal scientific

practice, and so they also have implications for what it means to say there is an abductively meaningful analogy between children and scientists. Paying attention to normativity and situatedness in both children and scientific research can help us understand how both solve the problem of induction. But attention to the same also reveals some important disanalogies between children and scientists that are not fully consistent with the original formulation of CATH.

Thus, since the fairest way to evaluate CATH is within a frame that makes children and scientists as similar as possible, and taking into account scientific evidence produced by research inspired by the original formulation of CATH, my talk will explore how children and scientists both "solve the problem of induction" using a frame that centers the normativity and the situatedness of learning in children and scientific inquiry. I will try to show that, within this frame, and in light of the scientific evidence from developmental psychology, there are important analogies and disanalogies which, when evaluated as a philosophical package, helpus formulate a more substantive, and thus substantially renovated, version of the child as scientist hypothesis.

The Logic and Semantics of Approximation in Models and their Solutions

Nicolas Fillion

To a large extent, the history of applied mathematics is one of becoming increasingly more proficient at using inexact mathematics in scientific endeavors. It is thus no surprise that philosophers of science have become more concerned with idealization, approximation, and solutions obtained via perturbation theory or numerical methods. Yet, at the formal level, until a substantive account of the notion of approximate truth is developed, many of the general claims about the inferential and representational role of inexact mathematics remain "just so much mumbo-jumbo," to use Laudan's phrase. Of course, this claim is not meant to dismiss the undeniable value of informal or semi-formal accounts of the way in which approximate truth operates in scientific methodology. Rather, the point is that to make the sort of general claims that would be required to ground adaptations of the mapping account or the inferential account of representation in a way that incorporates the realities of inexact mathematics and the hard-earned wisdom developed by applied mathematicians, a more formal account of approximate truth would be needed. In the mapping and inferentialist accounts of representation, first-order logic and its underlying satisfaction semantics remains the guiding paradigm. This being the case, based on the formal work of applied mathematicians in perturbation theory and numerical analysis, this talk systematically examines the analogies and disanalogies between truth semantics and approximate truth semantics, thereby showing that the two styles of semantics have radically different modi operandi.

Section 1 of the paper highlights the semantic distinction between classificatory and quantitative concepts, in the spirit of the measurement theory pioneered by Scott & Suppes. Section 2 argues that the notion of satisfaction and validity relevant to inexact representation cannot be treated using the standard syntactic (schematic) tools of formal logic, a key point for handling quantitative concepts. Section 3 abstracts the main operational concepts used to semantically assess inexact representation and inferences from the methods deployed in perturbation theory and numerical methods. Finally, section 4 isolates the conditioning of model equations as a hybrid concept combining the relevant internalist and externalist features that effectively enables scientists to correctly assess their inexact representations and inferences.

Isomorphism is Not Representation

Patrick Fraser

Isomorphism is Not Representation

Section: (a) General Philosophy of Science.

Keywords: Scientific representation; Isomorphism; Category theory; Models; Abstract and concrete entities.

Short Abstract. It is often held that a theoretical model represents a target system by sitting in some isomorphic relation to that target systems. I clarify the necessary ontological commitments one must have in order to sustain an isomorphism account of representation. I then show where claims about representation via isomorphism fail if one is unwilling to adopt these strong commitments. In particular, the viability of a morphism account of representation depends on one's commitments to the distinction between abstract and concrete entities. If concrete entities differ from abstract entities, then there cannot exist a mathematical category which contains both concrete and abstract entities, whence there can be no well-formed isomorphisms relating the two; only by representing concrete entities abstractly can such a morphism be defined, making the account of representation circular.

Extended Abstract. Theoretical models are often taken to be the most common vehicle through which scientists represent target systems in the world. But what sort of relationship must obtain between a theoretical model and its intended target system in order to warrant claims about its representational capacities? Often, the existence of some formal relation—typically expressed as a kind of isomorphism—between the abstract structure of the model and the concrete structure of the target system has been taken to ground claims about representation. Such isomorphism accounts of representation have been objected to on the grounds that they are inadequate for capturing actual scientific practice. In particular, it has been argued that the models may acquire their representational capacity in virtue of their pragmatic inferential utility, or as mediating instruments. In short, usual objections to isomorphism accounts of representation argue that such a view is too narrow to adequately express what is meant by representation in its full generality. Here, I follow a different tack; I demonstrate that even in the narrow setting in which the existence of an isomorphism might plausible account for some representational relation, such a notion of representation cannot be formally well-posed without becoming circular.

The argument I offer in favour of this thesis is as follows: target systems are concrete entities, whereas theoretical models are abstract entities. That is, a target system is a tangible thing that may or may not physically exist in the world, whereas a theoretical model is strictly a formal construction. Maintaining that abstract entities are different kinds of things from concrete entities, the structure of a model is identical to its abstract mathematical structure, while the structure of a target system is *not* identical to any abstract structure. Rather, it is through a relation of representation that we may attribute abstract formal structure to a target system.

Can such a relation literally take the form of an isomorphism? At the highest level of generality, isomorphisms are defined categorically. Mathematically, a category is a class of objects related by arrows that may be (associatively) composed together. An isomorphism is an arrow between objects in some category which has certain properties (it is epic and monic). As such, the relata that may be related by an isomorphism must always be objects that reside in the same category.

To say that a model represents a target system in virtue of their being isomorphic, we readily see that both the model (an abstract entity) and the target system (a concrete entity) must be objects in some common category. If one takes concrete entities to differ from abstract entities, however, then there is some constitutive feature that is possessed by tokens of one but not the other. Hence, if there is to be a category which includes both abstract and concrete entities as its objects, such a category must be forgetful of such constitutive properties. But identifying which features of a target system *qua* concrete entity are *not* salient to its formal structure (i.e. determining what may be forgotten) requires that one already know what the salient abstract structural features of the target system are. This in turn amounts to having an understanding of what sorts of abstract mathematical structures are adequate for representing the target system in the first place. Hence, in order to even write down the category within which an isomorphism between target system and model would be well-formed, one must already deploy a fully-fledged account of scientific representation to begin with. In short, isomorphisms only establish representational relations between target systems and their models if one already has a robust, detailed account of what it is in virtue of which the model is capable of representing the target system in the first place. Thus, one all of the representational considerations must be developed and employed before an isomorphism can even be made precise. As such, isomorphisms cannot warrant claims of representational capacities. Further connections to functorial relations between categories and natural transformations between functors are also discussed in the context of scientific representation.

The interplay of external and internal semiotics of domain-specific scientific theories

Alexander Gabovich and Volodymyr Kuznetsov

We suppose that those who are reading this contribution are familiar with the traditional reconstruction of domain-specific scientific theories (DSSTs) as logically ordered and static systems of statements about their domains. The latter split into separate realities with their attributes. However, any DSST (e.g., celestial mechanics, theories of superconductivity, theories of elementary particles) is an ever-improving tool for acquiring new knowledge. It means that a more realistic reconstruction is a varied polysystem. Its interacting and changing subsystems (SS) perform specific functions in the complex process of obtaining and testing the new knowledge. According to the modified structure-nominative reconstruction, there are many SSs in DSSTs [2019; 2021].

At any moment in the history of a DSST, its *ontic* subsystem contains the notions of realities and their attributes (properties, interactions, states, and processes) in question (for instance, "planet"/Mars, "mass"; "conductor"/ "temperature", "resistance"; "particle"/ "proton", "spin"). The content of the SS concerned varies and becomes more complex due to both the experimental progress and theoretical development.

Main components of the *model* SS are models of a different kind. They represent those attributes of the realities that are important for their study under certain experimental and theoretical settings. For instance, experimentally tested models of elementary particles appear while using accelerators testing higher and higher particle energies. In a first approximation, there are *verbal/visual/intuitive*, *empirically informative,* and *mathematical* models that are integrated in the appropriate *subsubsystems* of the model SS.

A language SS contains and orders languages that are used by the DSST. Each SS has a specific net of languages that describe its components.

A nomic SS contains formulations of laws, axioms, and postulates, which represent such theoretical attributes as regularities of realities from its domain, as well as the principles of organizing and changing the theory itself.

Other SS are *definitional* (formal and informal, full and partial definitions both of the realities/attributes from the theory domain and components of the theory); *ordering* (deductive, inductive, abductive, taxonomic and the like means of assembling other subsystems of the theory); *problem* (problems, questions and tasks that are formulated and solved by the theory); *operational* (operations both with the components of the theory and with the theory itself); *procedural* (procedures as rules for performing actions); *evaluative* (evaluations of components and the theory); *hypothetical* (plausible hypotheses generated by the theory); *heuristic* (useful but not well justified heuristic considerations); *approximate* (approximations of the theory and its components) and *connecting* (connections of both subsystems and their internal components) subsystems.

All components of subsystems mentioned above have general and specific names. It gives reasons for separating a *denominative SS*. Indeed, thinking about the notions of domain realities and their attributes is impossible without using distinct kinds of reality/attribute names. The *ontic subsubsystem* of the *denominative SS* of the theory includes various kinds of names (labels, designations, acronyms, terms, symbols, diagrams, schemes, tables, and the like), which represent the domain of the theory. The ontic names of realities and their attributes are borrowed from the national natural language and the universal *physical lingua franca* [2020].

The theoretical subsubsystem of the denominative SS includes similar means of naming the theory's internal components. Sometimes the same name denotes both the reality/attribute and the corresponding component of the theory. An example is a symbol ***E**,* which represents the actual electric component of the electromagnetic field and the corresponding vector function in the Maxwellian classical electrodynamics.

There are many ways to introduce and use new domain names. As an illustration, let's take the simplest case associated with the discovery of a new property *P* (its existence is taken for granted) of known realities. Let's denote its name as *N*(*P*). The obvious task of theorists is to theoretically calculate the values of *P* and then compare them with experimentally measured values. To do this, one needs to modify and use some internal components of DSST.

At the first stage, theorists should name the new problem *N*(*PR*) of calculating the values of the considered property *P*, and then formulate the corresponding *PR*(*P*) problem. In the absence of suitable models in the DSST, theorists should name the model *N*(*M*) and construct a new model *M*(*PR*) in terms of which it is promising to solve *PR*(*P*). All these components are specific in the sense that they refer to *N*(*P*).

The next step is attempts to resolve the *PR* by the existing structures from the operational SS or the name of *N*(*OP*) and the construction of new operational means of the *OP*. This event occurred in the history of quantum mechanics and is associated with the usage of matrices as new *OP*s. In the case of an approximate coincidence of the obtained numerical solution of the *RE*(*PR*), i.e., the calculated values of *P*, with the measured values, the *PR*(*P*) is conditionally/temporarily solved. As a rule, a more accurate measurement of property values creates the problem of reformulating the *PR*, as was the case with the so-called Lamb shift of the energy levels of atoms under the action of virtual particles emerging from the vacuum.

Otherwise, the cycle $P \rightarrow N(P) \rightarrow N(PR) \rightarrow PR \rightarrow N(M) \rightarrow M(PR) \rightarrow N(RE(M(PR))) \rightarrow RE(M(PR))$ stimulates the construction of a new theory with a new nomic SS. Here *RE*(*M*(*PR*)) denotes *PR* solution processes in terms of *M*(*PR*).

Thus, in modern physics, existential statements about realities should be supplemented by a theoretical calculation of their values and a comparison with measured values. These calculations are performed by working with changes to the theory's internal components, as well as the internal domain and theory names. It is important to note that while some theoretical names (for example, model names and calculated values of reality properties) indicate certain pieces of knowledge about realities, their domain names are quite arbitrary and might be independent of the nature of the named realities.

Philosophical Foundations of Meta-anthropology

Ilya Garber

Section D) Philosophy of Social Sciences

Philosophical Foundations of Meta-anthropology

Extended abstract

The purpose of the study is to present the philosophical foundations of meta-anthropology, the anthropology of anthropology, a special version of philosophical anthropology, an independent scientific discipline that develops in the image and likeness of other meta-sciences - meta-mathematics and meta-logic, meta-ethics, and meta-sociology. It is shown that the initial stages of the creation of meta-anthropology are closely related to philosophy. The term 'meta-anthropology' was coined by the American anthropologist and philosopher David Bidney. He tried to highlight the metaphysical aspect of anthropology and identified three approaches to its application: pre-scientific, post-scientific, and super-scientific. In addition, he outlined meta-anthropological boundaries (meta-anthropology deals with the ontological roots of anthropology) and called for the formulation of fundamental ontological postulates that underlie anthropology. 29 years after his first publication (Bidney, 1949), there was a theoretical breakthrough in meta-anthropology associated with S. Lee Seaton & Karen Ann Watson-Gegeo. As a theoretical framework for studying the schools of anthropological theory, they chose the forgotten approach of the pragmatic philosopher Stephen Pepper - his world hypotheses and the root metaphor model. This helped them to describe meta-anthropology on the basis of a well-thought-out system of key definitions, postulates, and criteria for critical assessment of theories of culture of two types - concrete anthropological and general philosophical. As a result, it was possible to create and substantiate the classification of theories of culture according to Pepper's four root metaphors:

formism/structuralism - culture configuration, social structure, cognitive structures; formism/functionalism - functionalism, social organization; contextualism/culture history - historical particularism, cultural historicism; contextualism/diffusionism - Kulturkreis, heliocentrism, multicentrism; organicism/systems theory - cybernetic systems, information systems; organicism/social evolution - British social evolutionism, American social evolutionism; mechanism/cultural materialism - technoeconomics, technoenergism. The work of Lee Seaton and Karen Watson-Gegeo paved the way for various generalizations, from the obvious - using Lakoff's theory of conceptual metaphors or the transition from Pepper's theory to more popular ones, to fundamentally different possibilities, but they were never realized. After 31 years, the torch has moved from the philosophy-based works to the empirically oriented approach of Eike Hinz (2009), building a normative model and searching for constructive criteria and problem frames in meta-anthropology. Besides anthropological theories, Hinz is considering problems related to education (elaboration curricula and academic programs, publishing textbooks, teaching anthropology, informing, and educating the public), ethical issues and the norms of human conduct for doing research (support for the people being studied and concerned, supporting their abilities, and increasing autonomy and self-determination, building the identity and self-respect of the people concerned, full participation and comprehensive information, existential welfare, and mental well-being of the local population), feedbacks between anthropology and neighboring empirical sciences, organization and stages of anthropological work (documentation, analysis, and critical assessment of cultures and societies), the inclusion of anthropology and its role in solving world problems such as reconciliation and peace-making, human rights, government policy, and survival of mankind.

This approach is consistent with Zeitgeist, but narrows the philosophical basis of meta-anthropology and makes it difficult to transform anthropology due to the transition of anthropologists from thinking/thought to meta-thinking/meta-thought. To improve the situation, the author proposes some possible prospects for the development of meta-anthropology, borrowed from disciplines adjacent to anthropology (Valentin Turchin's meta-transitional methodology, George Ritzer's metatheorizing, Steven Wallis's integrative propositional analysis, Kristen Madsen's systematology, Paul Meehl's and David Faust's cliometric approach). Finally, limitations of the study and different versions of meta-anthropology are considered.

Keywords

Meta-anthropology, world hypotheses, root metaphor

Short abstract

The purpose of the study is to present the philosophical foundations of meta-anthropology. D. Bidney tried to highlight the metaphysical aspect of anthropology. S.L. Seaton & K.A. Watson-Gegeo used S. Pepper's model as a theoretical framework. This helped them to build a system of key definitions, postulates, and criteria. Their work paved the way for various generalizations, but they were never realized. The baton has moved to the empirically oriented approach (E. Hinz). Some prospects for the development of meta-anthropology are proposed. Limitations of the study and different versions of meta-anthropology are considered.

Analogical inference Bayesian style 2.0

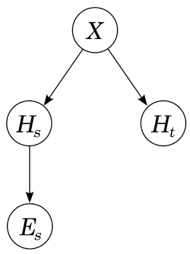
Alexander Gebharter and Barbara Osimani

f) Formal Philosophy of Science and Philosophy of Mathematics

Analogical inference Bayesian style 2.0

Keywords: analogical inference, Bayesian networks, confirmation

Scientists often rely on analogical inference. For example: Before a newly developed antiviral compound is tested on humans, it is tested on a suitable model organism such as rats. The evidence collected in the rat study is then used to provide a first evaluation of the hypothesis about the antiviral compound's effectiveness in humans. In this talk we explore the Bayesian model Dardashti, Hartmann, Thébault, and Winsberg (2019) developed as a general model for analogical inference. The aim of such a model is to spell out plausible assumptions which, if met, allow for analysing analogical inference in terms of Bayesian updating. Dardashti et al. propose that a Bayesian network with the following structure that satisfies the following constraints can do the job:



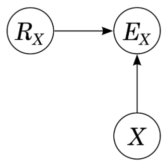
$$0 < Pr(\mathrm{X}) < 1 \tag{1}$$

$$Pr(\mathrm{H}_i\,|\,\mathrm{X}) > Pr(\mathrm{H}_i\,|\,\bar{\mathrm{X}}) \tag{2}$$

$$Pr(\mathrm{E}_s\,|\,\mathrm{H}_s) > Pr(\mathrm{E}_s\,|\,\bar{\mathrm{H}}_s) \tag{3}$$

s stands for the source system (rats) and t for the target system (humans). E_s models the evidence and H_i (with $i \in \{s, t\}$) the hypotheses about the antiviral compound's effectiveness. X models the structural similarity between the two systems under consideration. Upper case letters X in italics stand for binary variables and the non-italicised versions $\mathrm{X}, \bar{\mathrm{X}}$ for their positive/negative instantiations.

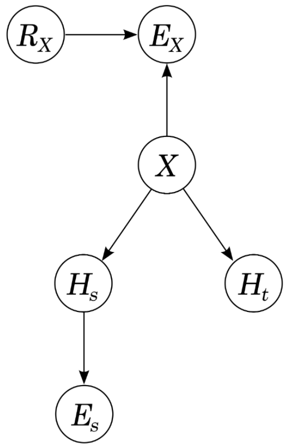
Equation 1 says that one should not assign extreme probabilities to the structural similarity X a priori. Equation 2 expresses the idea that the structural similarity should have a positive probabilistic impact on both hypotheses $\mathrm{H}_s, \mathrm{H}_t$. Finally, Equation 3 reflects the assumption that E_s is indeed (positive) evidence for H_s. All three assumptions are plausible.[1] Dardashti et al. (2019) show that these assumptions together indeed guarantee that E_s confirms H_t and, thus, that analogical inference reduces to Bayesian updating. In our talk we investigate how their model performs when varying the degree of certainty about the similarity between the source and the target system. This can be modelled on the basis of a reliability model (Bovens & Hartmann, 2003) which is a Bayesian network with the following structure satisfying the following constraints:



X	R_X	$Pr(\mathrm{E}_X\,|\,X, R_X)$
0	0	a
0	1	0
1	0	a
1	1	1

[1] For lack of space and since it can be solved easily we bracket a problem with Equation 2.

As before, X models the similarity of the two systems, E_X independent evidence for this similarity, and R_X the reliability of this evidence's source. One's certainty about the similarity X of s, t can now be modelled by assigning a prior probability to R_X and then conditioning on E_X. After combining the reliability model with Dardashti et al.'s (2019) model for analogical inference, we arrive at the following Bayesian network:



Based on the assumptions made so far we can make the following observations:

O1: $c(\mathrm{H}_t; \mathrm{E}_s) > 0$.[2]

O2: If $Pr(\mathrm{R}_X) = 1$, then $c(\mathrm{H}_t; \mathrm{E}_s \,|\, \mathrm{E}_X) = 0$.

O3: For some distributions Pr: $\Delta Pr(\mathrm{R}_X) > 0$ and $\Delta c(\mathrm{H}_t; \mathrm{E}_s \,|\, \mathrm{E}_X) < 0$.

All three observations are problematic. O1 means that E_s confirms H_t even if no evidence for the structural similarity X of s, t is considered. However, analogical inference should only be possible after having plausible reasons for assuming such a structural similarity. O2 says that having perfect evidence about the structural similarity results in no confirmatory impact of E_s on H_t at all. Finally, O3 tells us that sometimes an increase in certainty about the similarity results in a decrease of confirmatory impact of E_s on H_t, which stands in conflict with scientific practice. Here is an exemplary distribution Pr instantiating O3:

$Pr(\mathrm{E}_s \,|\, \mathrm{H}_s) = 0.9$

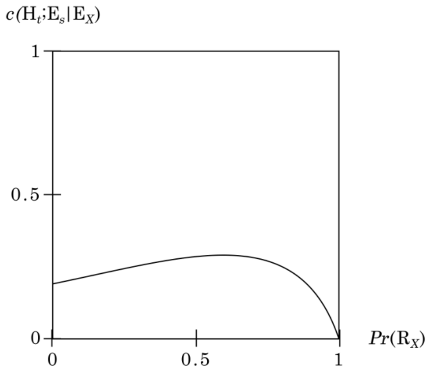
$Pr(\mathrm{E}_s \,|\, \bar{\mathrm{H}}_s) = 0.1$

$Pr(\mathrm{H}_i \,|\, \mathrm{X}) = 0.9$

$Pr(\mathrm{H}_i \,|\, \bar{\mathrm{X}}) = 0.1$

$a = 0.5$

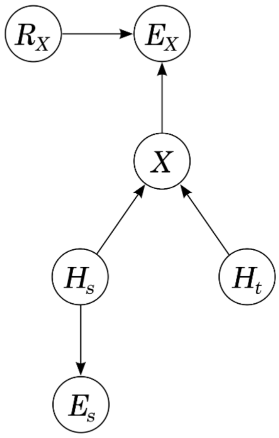
$Pr(\mathrm{X}) = 0.1$



---

[2] $c(\mathrm{H}_t; \mathrm{E}_s)$ is the ordinary Bayesian difference measure defined as $Pr(\mathrm{H}_t \,|\, \mathrm{E}_s) - Pr(\mathrm{H}_t)$.

We propose to modify the original Bayesian network's structure and to replace the original assumptions expressed by Equations 1-3 by the following constraints as an alternative Bayesian model for analogical inference:



$$0 < Pr(\mathrm{H}_i) < 1 \tag{4}$$

$$x_{\mathrm{H}} = x_{\bar{\bar{\mathrm{H}}}} > x_{\bar{\mathrm{H}}} \tag{5}$$

$$\begin{aligned} x_{\mathrm{H}} &= Pr(\mathrm{X}\,|\,\mathrm{H}_s, \mathrm{H}_t) \\ x_{\bar{\mathrm{H}}} = Pr(\mathrm{X}\,|\,\mathrm{H}_s, \bar{\mathrm{H}}_t) &= Pr(\mathrm{X}\,|\,\bar{\mathrm{H}}_s, \mathrm{H}_t) \\ x_{\bar{\bar{\mathrm{H}}}} &= Pr(\mathrm{X}\,|\,\bar{\mathrm{H}}_s, \bar{\mathrm{H}}_t) \end{aligned} \tag{6}$$

Equation 4 says that the prior probabilities one assigns to the hypotheses H_i (with $i \in \{s, t\}$) should not be extreme. Equations 5 and 6 together characterise the mechanism underlying the analogical inference: If s, t indeed share relevant structural features, then the two hypotheses $\mathrm{H}_s, \mathrm{H}_t$ are more likely to be true or false together (Equation 5). In addition, an actual structural similarity X does not probabilistically discriminate between one hypothesis being true and the other one being false (Equation 6).

Given the new structure and Equations 4-5, we can make the following observations about our modified model:

O4: If $Pr(\mathrm{R}_X) > 0$, then $c(\mathrm{H}_t; \mathrm{E}_s\,|\,\mathrm{E}_X) > 0$.

O5: $c(\mathrm{H}_t; \mathrm{E}_s) = 0$.

O6: If $Pr(\mathrm{R}_X) = 1$, then $c(\mathrm{H}_t; \mathrm{E}_s\,|\,\mathrm{E}_X) = max$.

O7: If $\Delta Pr(\mathrm{R}_X) > 0$, then $\Delta c(\mathrm{H}_t; \mathrm{E}_s\,|\,\mathrm{E}_X) > 0$.

O4 tells us that the model still allows to analyse analogical inference in terms of Bayesian updating. If Equations 4-6 are satisfied and one has some evidence for a structural similarity of s, t, then E_s indeed confirms H_t. O5 fixes the problem expressed by O1, O6 the one expressed by O2, and O7 the one expressed by O3. O5 says that if no evidence about the structural similarity is considered, there is indeed no analogical confirmation. O7 shows that an increase in certainty about the structural similarity goes always hand in hand with an increase in confirmatory impact E_s has on H_t. This corresponds to scientific practice. Take the rat case study from above as an example: There exist specific breeding programs aiming at making the immune system of rats more similar to the human immune system when it comes to the response to certain types of antiviral compounds. The goal is, of course, to create even better model organisms. The more similar the model organism becomes, the more impact the findings in the rat study do have on the corresponding hypothesis about humans. Finally, O6 is a direct consequence of O7. It tells us that having absolute certainty about the structural similarity results in the maximum confirmatory impact E_s can have on H_t.

References

Bovens, L., & Hartmann, S. (2003). *Bayesian epistemology*. Oxford University Press.

Dardashti, R., Hartmann, S., Thébault, K., & Winsberg, E. (2019). Hawking radiation and analogue experiments: A Bayesian analysis. *Studies in History and Philosophy of Modern Physics Part B*.

Mere prediction without understanding?

Lilia Gurova

Section: a) General Philosophy of Science

Title: Mere prediction without understanding?

Keywords: prediction and understanding, understanding theories, understanding phenomena

Short abstract

The discussions about the connection between prediction and understanding are surrounded by controversies, which are visible even in the most elaborated accounts of scientific understanding. So, de Regt (2017) on the one hand argues that "prediction turns out to be impossible without understanding" but on the other hand acknowledges the existence of "mere prediction without understanding". In this paper the alleged controversy is put under scrutiny to show that "no prediction without understanding" is a better starter, having implications that comply with the way prediction and understanding are seen in some areas of behavioral sciences.

Extended abstract

In one of the most elaborated accounts of scientific understanding, de Regt (2017) makes a clear distinction between "understanding a theory" and "understanding a phenomenon". A theory is understood by a scientists, or as de Regt put it, it is "intelligible" to him/her, if (s)he could use this theory to build descriptive, explanatory and predictive models of various phenomena. The latter, he argues, is impossible without having an intelligible theory, i.e. a theory, which the scientist understands. This is the basis of de Regt's claim that prediction is impossible without understanding (a theory). On the other hand, for de Regt, having an intelligible theory is necessary but not sufficient for understanding a particular phenomenon. To achieve such an understanding one should build an explanation based on an intelligible theory, i.e. according to de Regt, the understanding of a phenomenon is always an explanatory understanding. This leads him to a second claim that it is possible to predict phenomena without understanding them insofar as it is possible to predict without having an explanation of what is predicted. The views that "prediction turns out to be impossible without understanding" and that "mere prediction without understanding" exists (De Regt 2017, p. 107) seem incompatible but in fact they are not because the first view refers to understanding a theory and the second one refers to understanding a phenomenon (see Findl & Suárez, 2021 for a similar reading of this apparent contradiction). But is the distinction between understanding a theory and understanding a phenomenon justifiable? To see the problems, which such a distinction creates, one should turn to disciplines where understanding of a phenomenon is often gained in the absence of intelligible theories or data-driven descriptive models. Such cases are probably rare in natural science but they are typical for behavioral research where correlational studies play an important role. To see how mere correlations, which neither explain nor describe the correlated phenomena nonetheless enhance our understanding by allowing useful predictions, let's consider two examples (Yarkoni & Westfall, 2017).

Example 1: Applying multi-voxel pattern analysis (MVPA) to fMRI data recorded during a face recognition task, Rissman, Greely and Wagner (2010) have found that this data successfully predict the participants' subjective experience of whether a shown face has or has not been seen by them before. They have also found that the data cannot predict whether the face shown is objectively new to the participant in the experiment or (s)he has already seen it. The main findings of this study are based on a correlation obtained between two sets of data, which allows to predict the one set given the other. This prediction is neither based on an intelligible theory nor is it based on a model describing what underlies the established correlation. Nonetheless, the correlation allowing the specified prediction enhances our understanding of human memory as it shakes the widespread belief that our brains store objective traces of the actual events to which we have been exposed.

Example 2: Using linear and logistic regression models applied to data obtained from over 58000 Facebook (FB) profile owners, Kosinski, Stillwell and Graepel (2013) have found that the FB users' "likes" successfully predict a wide range of personal characteristics, including the personality traits of participants in the study. In this case, too, the main finding is based on a correlation between two sets of data, allowing to predict the one set given the other. Again, the prediction is not inferred from an intelligible theory or from a model describing the dynamics of the two domains, which turned to be connected. Nonetheless, the generated prediction is associated with an increase in our understanding of human personality and the way it determines human behavior as it shows that the real personality traits could be revealed by samples of behavior, which is intended to hide any negative aspects of the personality of those who exhibit such behavior.

The analysis of these examples suggests that:

(a) predictions seem to carry understanding about predicted phenomena even in the absence of explanations, intelligible theories or descriptive models of the data which make the predictions possible;

(b) in view of (a), one can argue that "no prediction without understanding" applies not only to the understanding of theories, as de Regt insists, but also to the understanding of phenomena;

(c) the conclusion (b) eventually questions the appropriateness of distinguishing between "understanding a theory" and "understanding a phenomenon".

References

De Regt, H. W. (2017). *Understanding scientific understanding*. Oxford: Oxford University Press.

Findl, J., & Suárez, J. (2021). Descriptive understanding and prediction in COVID-19 modelling. *History and philosophy of the life sciences*, *43*(4), 1-31.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the national academy of sciences*, *110*(15), 5802-5805.

Rissman, J., Greely, H. T., & Wagner, A. D. (2010). Detecting individual memories through the neural decoding of memory states and past experience. *Proceedings of the National Academy of Sciences*, *107*(21), 9849-9854.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100-1122.

Benefits and Challenges of using Qualitative Methods in Empirical Philosophy of Science

Nora Hangel

Talk proposal for the fourth conference of the East European Network for Philosophy of Science for the University of Tartu, August 17-19, 2022

To be considered for either of the sections:

e) History, Philosophy and Social Studies of Science or

d) Philosophy of Social Sciences

Title: Benefits and Challenges of using Qualitative Methods in Empirical Philosophy of Science

Keywords: scientific practice, descriptive values, normative values, boundary work, qualitative methods

Abstract: 100 words

In collaborative research, scientists' abilities to communicate clear accounts of criteria, norms, and standards used to evaluate empirical evidence are becoming crucially important for the functioning of science itself. Qualitative methods provide insights into concrete scientific practice and offer a richer understanding of how scientific reasoning and collaborative processes contribute to generating knowledge. I will present successes and benefits of qualitative methods in empirical philosophy of science and address scientific challenges: The subjectivity of qualitative methods and its relation to philosophical claims striving for universality; Discuss how with qualitative methods descriptive and normative values in scientific reasoning can be studied.

Abstract: (654 words; 811 including references)

Despite its promise to inform philosophical theory empirically and its considerable successes, qualitative methods have not been considered for philosophy of science practice as other methods e.g. historical case studies or quantitative methods e.g. agent-based modeling. The talk analyses scientific challenges, boundary problems, and inquires why qualitative methods have not yet received the boost in empirical philosophy of science they would deserve.

Several philosophers have already experimented with combining historical and quantitative methods (e.g., Frey & Šešelja, 2018). Yet, quantitative insights can be fruitfully augmented by more detailed qualitative studies. While quantitative methods can provide strong understanding of correlative and causative relationships between variables, often qualitative methods provide insights into concrete, scientific practice and offer a richer understanding of how scientific reasoning and collaborative processes contribute to generating knowledge. When philosophers like MacLeod & Nersessian (2016) studied interdisciplinary research teams, they focused on how these diverse groups of modelers and biologists adapt to the absence of shared disciplinary problems and norms and instead develop their alternative practices of problem solving. Specifically, in collaborative research, scientists' abilities to communicate clear accounts of the criteria, norms, and standards used to evaluate empirical evidence are becoming crucially important for the functioning of science itself (MacLeod & Nersessian, 2016). As collaborators in interdisciplinary contexts, scientists need conceptual tools to express, explain, justify, critique, and evaluate their methodological standards and engage in a meta-discourse about epistemic and social values in research to different audiences. To communicate the scope of their results as well as their own accountability, researchers often reflect on the potential tensions between social and cognitive/methodological criteria for decisions under uncertainty (Schickore & Hangel, 2019). When philosophers utilize qualitative methods, their usage differs significantly from sociologists. For instance, sociologists like H. Collins (2019) focus on the dynamics of collectivites in generating knowledge, such as how socialization into scientific communities is a disciplinary process for individuals.

In contrast, a philosophical adaptation of qualitative empirical philosophy of science aligns with Helen Longino's approach (2002) when arguing against the dichotomy between social and rational processes of belief formation. This approach retains keen interest in the *content* of such dynamics and the *quality* of the reasoning. We do not need to abandon epistemological questions but instead rearticulate them for non-idealized subjects by localizing and contextualizing justification and epistemic acceptability, recognizing the interdependence of cognitive agents when generating knowledge and also being aware that there is a plurality of knowledge. She describes epistemic acceptability in terms of procedural norms: having survived critical scrutiny from as many standpoints as possible, "uptake of criticism, public standards, and tempered equality of intellectual authority" (Longino, 2002, 135). Longino has been criticized e.g. by Solomon (1994) as overtaxing on the scientific community. However, if we apply Longinos criteria for knowledge not only to the scientific community at large but at the collaborative working group, the benefit becomes clear: Individual agency of knowledge becomes the (interpersonal) interdependence of cognitive agency that contributes to knowledge. Justificatory processes of credentialing include discursive interactions among researchers, criticisms from different perspectives, and examinations and evaluations of descriptive and normative implications are actual practice in every collaboration, experimental or otherwise. By reclaiming social practice as the object of philosophical inquiry, Longino emphasizes knowledge as concurrently social and cognitive (Longino, 2002, 204). I will argue, that with qualitative methods we can utilize Longino's theory for empirically grounded philosophical analysis.

Thus, the talk will present successes and benefits of using qualitative methods in empirical philosophy of science and address scientific challenges: first, the subjectivity of qualitative methods and its relation to philosophical claims that strive for universality; second, I will discuss how with qualitative methods we can incorporate descriptive and normative components. To study scientific reasoning, I draw on social epistemologists like S. Goldberg, who argues to access knowledge practices normatively presupposes an accurate characterization, which again relies on description (Goldberg, 2020, 417). I will propose how with qualitative methods we can support these aims.

References:

Collins, H. (2019). Forms of Life: The Method and Meaning of Sociology. Cambridge, MA: The MIT Press.

Frey, D., & Šešelja, D. (2018). What Is the Epistemic Function of Highly Idealized Agent-Based Models of Scientific Inquiry? Philosophy of the Social Sciences, 48(4), 407–433. https://doi.org/10.1177/0048393118767085

Goldberg, S. C. (2020). Social Epistemology: Descriptive and Normative. In M. Fricker, P. J. Graham, D. Henderson, N. J. L. L. Pedersen, & J. Wyatt (Eds.), Routledge handbooks in philosophy. The Routledge handbook of social epistemology (1st ed., pp. 417–424). London: Routledge.

Longino, H. E. (2002). The Fate of Knowledge. Princeton, NJ [u.a.]: Princeton Univ. Press.

MacLeod, M., & Nersessian, N. J. (2016). Interdisciplinary Problem-Solving: Emerging modes in integrative systems biology. European Journal for Philosophy of Science, 6(3), 401–418.

Schickore, J., & Hangel, N. (2019). "It might be this, it should be that…" uncertainty and doubt in day-to-day research practice. European Journal for Philosophy of Science, 9(2), 1. https://doi.org/10.1007/s13194-019-0253-9

Crossing Domains: The Role of Translation in Model Transfer

Catherine Herfeld

Short Abstract

The transfer of models has often been considered an essential prerequisite for progress in science. However, a possible source of tension such transfers confront is that, in order to be successfully applied, the model to be transferred from some source domain not only has to be novel, but also align with area-specific standards and theoretical frameworks already existing in the target domain. In this paper, I analyze how this tension is resolved in practice. In particular, I suggest that the successful spread of a scientific model involves a process of 'translation' (Kuhn 2013 [1977]). The goal of this paper is to further unpack what kind of knowledge this translation process requires in the case of cross-disciplinary model transfer. More specifically, I discuss two kinds of knowledge required to successfully 'translate' a model. I argue in line with Paul Humphreys that models may in principle be detachable from their theory of origin, which is why their application in the target domain does require knowledge of that domain. However, contrary to Humphreys, I argue that this process of detachment nevertheless requires knowledge about the domain it was constructed in. Without such knowledge, the scientist may not only be unable to engage with the template at the appropriate level of abstraction and consequently fail to modify its associated content. Such knowledge is also required to turn what Humphreys calls a theoretical template into an applicable model. I support my claims by discussing the specific case of transferring rational choice theories from mathematics into political science. The analysis partly explains why some mathematical models spread across different disciplines while others do not. Furthermore, I discuss the implications of my analysis for the design and organization of successful interdisciplinary research environments.

Long Abstract

Discussions in philosophy of science have emphasized the importance of a cross-domain transfer of scientific models as one of the central catalysts for scientific innovation and progress in a field. Existing philosophical analyses have long rested on the implicit assumption that a model does not have to undergo substantial changes in the transfer process. And indeed, as for example the case of 'economics imperialism' shows, model transfers can consist in applying a model originating in some source domain without any modification to phenomena in some target domain. However, model transfer can take different forms. This implies that in some cases, successful transfers presuppose a partial adaptation to, or even a full integration of, the transferred model with the knowledge already accepted in the target domain. As philosophers in some recent case studies have shown, modifying the model transferred is often indispensable for a successful cross-domain transfer (e.g., Herfeld and Lisciandra 2019). For example, models from engineering have been increasingly used in synthetic biology only after extensive modification, where they have called into

question a variety of basic principles of existing theoretical frameworks (e.g., Knuuttila and Loettgers 2016, 2014). In short, model transfer often includes modification of the unit of transfer to be successful.

One reason for why models have to undergo such modification processes to make them applicable in the target domain is that they can help overcoming a number of challenges that model transfer processes might confront. One such challenge arises out of the need for the model to bridge the gap between novelty a¬nd tradition. More specifically, in order to spread successfully across domains, the model originating in some source domain has to align with specific methodological and epistemic standards as well as accepted theoretical frameworks and concepts of the target domain. Economics as a discipline is a prime example in which attempts for model transfers from physics, sociology, and psychology have been manifold but where such transfers often have to overcome substantial barriers to entry or sometimes even fail completely because of the absence of such alignment (e.g., Bradley and Thébault 2019). As such, to better understand how model modification addresses this tension and thereby overcomes the challenge, both the conservation and adaptation aspects need to be studied.

In this paper, I analyze how this tension between conservation and adaptation via modification is resolved in social scientific practice. The analysis thereby addresses the more general question of how successful model transfer across different scientific domains can be explained. In particular, I suggest that resolving this tension involves a process of 'translation' (Kuhn 2013 [1977], Herfeld and Doehne 2018). The main goal of the paper is to further unpack what kind of knowledge such translations require. Roughly, I suggest that translation manifests as a two-directional modification process that balances conservation and adaptation in such a way that the application of the model is epistemically beneficial for the target domain. More specifically, translation of a model is successful when a balance between conservation and adaptation can be found by what Hasok Chang (2004) called 'epistemic iteration,' a process by means of which knowledge claims become progressively adjusted and refined (Chang 2004, Elliott 2012).

My analysis draws on Paul Humphrey's concepts of 'theoretical' and 'computational templates.' I analyze what Humphreys has labelled the "construction process" in the case of applying a novel theoretical model within and across scientific domains (Humphreys 2002, 2004). The concept of a construction process allows for systematically understanding the kinds of modifications a model undergoes when being transferred from the source domain into the target domain in order for it to be successfully applied to problems in the latter. According to Humphreys, formal templates may in principle be detachable from their domain of origin. As such, the construction process only requires target domain-specific knowledge. To transfer and ultimately apply such templates, scientists do not need to draw on knowledge of the source domain according to Humphreys. In short, there is "no need for vocabulary translations or for interdisciplinary knowledge" (2020, 7).

I argue in line with Humphreys that, first, the model that is newly introduced into a specific target domain becomes aligned with knowledge established in that domain. To do so, the translation of a model requires target domain-specific knowledge in order to select appropriate idealizations and abstractions to justify the modification procedure, interpret the template, and finally apply it to a specific problem in the target domain (Humphreys 2008, 174). Second, however, I argue in line with Kuhn that because large parts of the model are formulated in the language of the source domain, knowledge of that domain is equally required for successful modification. Knowledge of the source domain not only enables the scientist to engage with the template in the first place; without it, the scientist may not be able to engage with the template at the appropriate level of abstraction and to recognize the potential of a template for a specific problem in her domain. Knowledge of the source domain is also required for further specifications of the template to turn it into an applicable model, such as for selecting which concepts from the source domain are essential and needed for its application, what the scope of the template is, which feasibility constraints there are, etc. As such, both kinds of knowledge are necessary to cope with the aforementioned tension between novelty and tradition in a way that model transfer leads to epistemic iteration.

I support my claims by focusing on one of the most widely spread templates in the social sciences, namely rational choice theories. First, I further clarify Humphreys' categories of theoretical, computational, and trans-domain template by applying them to the transfer of rational choice models from mathematics to the social sciences. Second, I discuss the role that translation played in their transfer into political science in particular and the kind of knowledge required for such translation. My analysis offers a set of key insights into the conditions under which mathematical models disseminate across social scientific domains. Furthermore, I discuss the implications of those insights for the design and organization of successful interdisciplinary research environments.

References

Bradley, Seamus, Thébault, Karim (2019). Models on the Move: Migration and Imperialism. Studies in the History and Philosophy of Science: Part A, 77, 81–92.

Chang, Hasok (2004). Inventing Temperature: Measurement and Scientific Progress. New York, Oxford University Press.

Elliott, K.C. (2012). Epistemic and Methodological Iteration in Scientific Research. Studies in History and Philosophy of Science: Part A, 43 (2), 376–382.

Herfeld, Catherine, Doehne, Malte (2019). The Diffusion of Scientific Innovations: A Role Typology, Studies in History and Philosophy of Science Part A, 77, 64-80.

Herfeld, Catherine, Lisciandra, Chiara (2019) (eds.). Special Issue on 'Knowledge Transfer and Its Contexts'. Studies in History and Philosophy of Science: Part A 77; URL: https://www.sciencedirect.com/journal/studies-in-history-and-philosophy-of-science-part-a/special-issue/10WK61R4ZW9 Humphreys, Paul (2002). Computational Models, Philosophy of Science, 69 (S3), S1–S11.

Humphreys, Paul (2004). Extending Ourselves: Computational Science, Empiricism, and Scientific Method, Oxford: Oxford University Press.

Knuuttila, Tarja, Loettgers, Andrea (2020). Magnetized Memories: Analogies and Templates in Model Transfer, in: Holm, Sune, Serban, Maria (eds.): Philosophical Perspectives on the Engineering Approach in Biology: Living Machines?, London: Routledge.

Knuuttila, T., & Loettgers, A. (2016). Model Templates within and between Disciplines from Magnets to Gases – and Socio-Economic Systems. European Journal for the Philosophy of Science 6, 377–400.

Knuuttila, T., & Loettgers, A. (2014). Magnets, Spins, and Neurons: The Dissemination of Model Templates Across Disciplines. The Monist 97, 280–300.

Kuhn, Thomas (2013 [1977]). Objectivity, Value Judgment, and Theory Choice, in: Curd, Martin, Cover, J.A., Pincock, Christopher (eds.): Philosophy of Science: The Central Issues, 2nd ed., New York, London: W.W. Norton & Company; 94-110.

In silico methods – simulations or experiments? : Computational aspects of demarcation

Michal Hladky

Abstract for the EENPS 2022, University of Tartu, Estonia, August 17-19, 2022

In silico methods – simulations or experiments?

Computational aspects of demarcation

Sections

b) Philosophy of Natural Science

f) Formal Philosophy of Science and Philosophy of Mathematics

a) General Philosophy of Science

Keywords

Computer simulation, Experiments, In silico experiments, Demarcation, Neuroscience

Short abstract

The epistemic power of omnipresent computer simulations is often evaluated relative to experimental methods. A preliminary condition for such evaluations is the identification of properties which make computer simulations different from experiments. Several authors evoke formal or material features of computation in order to state the distinction. In silico experiments used in neuroscientific research are particularly pertinent for this debate as they exhibit features of computer simulations and of experiments. Through the analysis of in silico methods deployed in the Blue Brain Project (BBP), I demonstrate that neither formal, nor material aspects of computation are sufficient criteria for demarcation.

Abstract

The epistemic power of omnipresent computer simulations is often evaluated relative to experimental methods (Winsberg 2009; Parker 2009; Parke 2014; Roush 2017). A preliminary condition for such evaluations is the identification of properties which make computer simulations different from experiments.

In silico experiments, or more neutrally in silico methods, are deployed in major neuroscientific research endeavours – the Human Brain Project (HBP) and the Blue Brain Project (BBP). These methods exhibit features of both simulation and of experiments. Through a case study based on the results of BBP (Markram et al. 2015) and using notions and results of model theory, I evaluate formal and material aspects of computation that have been suggested to demarcate simulations from experiments. This formal approach allows for an analysis independent from any strong reliance on intentional states of researchers (goals such as hypothesis confirmation, discovery of causal relations, generation of surprising observations) or their metaphysical commitments.

First, I evaluate formal criteria revolving around abstract computation and deductive inference, that are often treated as indications that computer simulations are theoretical and not experimental (Beisbart 2012; 2018; Beisbart and Norton 2012; similar considerations exposed in Galison 1996). I show that deduction as demarcating criterion faces a series of obstacles which are increasingly difficult to overcome: i) the predicate mismatch between the languages of the sources and the targets; ii) emergent behaviour (Bedau 2008) and inhomogeneous reductions (Nagel [1970] 2008); iii) incompleteness of theories relative to partial sampling; iv) huge space of possibilities in case of stochastic simulations. Furthermore, I demonstrate that, contrary to Beisbart and Norton (2012), the BBP stochastic framework does not deploy a Monte Carlo style analysis to generate the results.

Second, I show that material aspects of computation as demarcating criterion can be considered problematic because of strong metaphysical assumptions about i) abstractness of computation; ii) causality (Massimi and Bhimji 2015; Guala 2002; Guala and Mittone 2005); iii) natural kinds; iv) identity (Beisbart 2018). The most general criterion – identity – has to be interpreted as type identity, in order to avoid implausible notion of experiments. Furthermore, without strong metaphysical assumptions about natural kinds, the distinction between source and target classes, often used to

characterised models and simulations, can be overcome by defining one broad experimental class that includes the computer systems and the targets (biological samples), effectively collapsing the sources and the targets into a single class.

Finally, I demonstrate that in silico methods used in BBP can be reconstructed as simulations and as experiments, effectively undermining several of the proposed demarcating criteria based on formal or material aspects of computation. This does not mean that the concepts of experiment and of simulation can be used interchangeably. Rather, it indicates the necessity to consider broader methodological aspects – type of inference, background assumptions and their justification – in order to distinguish between the two methods.

References

Babai, László. (1979) 2005. 'Monte-Carlo Algorithms in Graph Isomorphism Testing'.

Bedau, Mark A. 2008. 'Weak Emergence'. *Noûs* 31 (June): 375–99. https://doi.org/10.1111/0029-4624.31.s11.17.

Beisbart, Claus. 2012. 'How Can Computer Simulations Produce New Knowledge?' *Euro Jnl Phil Sci* 2 (3): 395–434. https://doi.org/10.1007/s13194-012-0049-7.

———. 2018. 'Are Computer Simulations Experiments? And If Not, How Are They Related to Each Other?' *European Journal for Philosophy of Science* 8 (2): 171–204. https://doi.org/10.1007/s13194-017-0181-5.

Beisbart, Claus, and John D. Norton. 2012. 'Why Monte Carlo Simulations Are Inferences and Not Experiments'. *International Studies in the Philosophy of Science* 26 (4): 403–22. https://doi.org/10.1080/02698595.2012.748497.

Boge, Florian J. 2018. 'Why Computer Simulations Are Not Inferences, and in What Sense They Are Experiments'. *European Journal for Philosophy of Science* 9 (1): 13. https://doi.org/10.1007/s13194-018-0239-z.

Button, Tim, and Sean Walsh. 2018. *Philosophy and Model Theory*. New product edition. NewYork, NY: Oxford University Press.

Fodor, Jerry. 2007. 'The Revenge of the Given'. In *Contemporary Debates in Philosophy of Mind*, edited by Brian P. McLaughlin and Jonathan D. Cohen, 105–16. Contemporary Debates in Philosophy. Malden, Mass: Blackwell Pub.

Frigg, Roman, and Julian Reiss. 2009. 'The Philosophy of Simulation: Hot New Issues or Same Old Stew?' *Synthese* 169 (3): 593–613. https://doi.org/10.1007/s11229-008-9438-z.

Galison, Peter. 1996. 'Computer Simulations and the Trading Zone'. In *The Disunity of Science: Boundaries, Contexts, and Power*, edited by Peter Galison and David J. Stump, 118–57. Writing Science. Stanford, Calif: Stanford University Press.

Gilbert, G. Nigel, and Klaus G. Troitzsch. (1999) 2005. *Simulation for the Social Scientist*. 2nd ed. Maidenhead, England ; New York, NY: Open University Press. http://www.modares.ac.ir/uploads/Agr.Oth.Lib.16.pdf.

Guala, Francesco. 2002. 'Models, Simulations, and Experiments'. In *Model-Based Reasoning*, edited by Lorenzo Magnani and Nancy J. Nersessian, 59–74. Boston, MA: Springer US. http://link.springer.com/10.1007/978-1-4615-0605-8_4.

Guala, Francesco, and Luigi Mittone. 2005. 'Experiments in Economics: External Validity and the Robustness of Phenomena'. *Journal of Economic Methodology* 12 (4): 495–515. https://doi.org/10.1080/13501780500342906.

Hammersley, John M., and David C. Handscomb. (1964) 1975. *Monte Carlo Methods*. Reprint. Methuen's Monographs on Applied Probability and Statistics. London: Methuen.

Hartmann, Stephan. 1996. 'The World as a Process: Simulations in the Natural and Social Sciences'. In *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View*, edited by Rainer Hegselmann, Uli Mueller, and Klaus G Troitzsch, 77–100. Dordrecht: Springer Netherlands. http://dx.doi.org/10.1007/978-94-015-8686-3.

Humphreys, Paul. 2009. 'The Philosophical Novelty of Computer Simulation Methods'. *Synthese* 169 (3): 615–26. https://doi.org/10.1007/s11229-008-9435-2.

James, F. 1980. 'Monte Carlo Theory and Practice'. *Reports on Progress in Physics* 43 (9): 1145–89. https://doi.org/10.1088/0034-4885/43/9/002.

Markram, Henry, Eilif Muller, Srikanth Ramaswamy, Michael W. Reimann, Marwan Abdellah, Carlos Aguado Sanchez, Anastasia Ailamaki, et al. 2015. 'Reconstruction and Simulation of Neocortical Microcircuitry'. *Cell* 163 (2): 456–92. https://doi.org/10.1016/j.cell.2015.09.029.

Massimi, Michela, and Wahid Bhimji. 2015. 'Computer Simulations and Experiments: The Case of the Higgs Boson'. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 51 (August): 71–81. https://doi.org/10.1016/j.shpsb.2015.06.003.

Morgan, Mary S. 2003. 'Experiments Without Material Intervention: Model Experiments, Virtual Experiments and Virtually Experiments'. In *The Philosophy of Scientific Experimentation*, edited by Hans Radder, 216–35. Pittsburgh, Pa: University of Pittsburgh Press. http://digital.library.pitt.edu/cgi-bin/t/text/text-idx?c=pittpress;cc=pittpress;view=toc;idno=31735047065523.

Nagel, Ernest. (1970) 2008. 'Issues in the Logic of Reductive Explanations'. In *Emergence*, edited by Mark A. Bedau and Paul Humphreys, 359–73. The MIT Press. http://mitpress.universitypressscholarship.com/view/10.7551/mitpress/9780262026215.001.0001/upso-9780262026215-chapter-19.

Parke, Emily C. 2014. 'Experiments, Simulations, and Epistemic Privilege'. *Philosophy of Science* 81 (4): 516–36. https://doi.org/10.1086/677956.

Parker, Wendy S. 2009. 'Does Matter Really Matter? Computer Simulations, Experiments, and Materiality'. *Synthese* 169 (3): 483–96. https://doi.org/10.1007/s11229-008-9434-3.

Robinson, Abraham. 1951. *On the Metamathematics of Algebra*. Studies in Logic and the Foundations of Mathematics. Elsevier Science.

Roush, Sherrilyn. 2017. 'The Epistemic Superiority of Experiment to Simulation'. *Synthese*, May. https://doi.org/10.1007/s11229-017-1431-y.

Weber, Marcel. 2014. 'Experimental Modeling in Biology: In Vivo Representation and Stand-Ins as Modeling Strategies'. *Philosophy of Science* 81 (5): 756–69. https://doi.org/10.1086/678257.

Winsberg, Eric. 2009. 'A Tale of Two Methods'. *Synthese* 169 (3): 575–92. https://doi.org/10.1007/s11229-008-9437-0.

The aesthetic value of scientific experiments

Milena Ivanova

The aesthetic value of scientific experiments

Abstract

I explore the aesthetic dimensions of scientific experimentation, addressing specifically the question how aesthetic features enter the construction, evaluation and reception of an experiment. I analyse which aspects of experiments are appreciated aesthetically in modern experiments identifying several contenders, from the ability of an experiment to uncover nature's beauty, to encapsulating original designs and human creativity. Following this analysis, I focus on the notion of beauty: what makes an experiment beautiful? Several common qualities are explored, from the simplicity and economy of the experiment, to the significance of the experimental results.

Scientific products have long been valued for their aesthetic features and compared to works of art. We hear that Einstein's relativity theory, the double helix structure of DNA and images of colliding particles are beautiful, and that just like works of art, they evoke in us aesthetic responses. Scientists themselves, like artists, are praised for their creativity, originality and aesthetic sensibility. Einstein famously argued that Michelson, who designed the famous experiment to measure the velocity of the Earth relative to the ether, was 'the artist in science', claiming that Michelson did not only care for devising a good experiment, but wanted his creations to be beautiful too.

In this paper I ask an underexplored question in philosophy: what makes scientific experiments aesthetically valuable? I start by presenting a three-fold way to think about the aesthetic value of scientific experiments. First, there is an obvious immediate answer to this question: scientific experiments reveal pleasing phenomena or pleasing set ups. This answer certainly captures something important about experiments, they can reveal to us nature's beauty and they can do so by utilising pleasing instruments, but I will argue that it is not a satisfactory answer. Let us consider an example to illustrate why this is so. Foucault's Pendulum experiment allows us to illustrate three important ways in which an experiment can be beautiful. The experiment was designed to demonstrate that the Earth rotates on its axis. In 1851 Léon Foucault hung a heavy brass weight from a long cable fixed to the inside of the dome of the Pantheon in Paris. When he set this pendulum in motion it swung slowly back and forth tracing lines in sand beneath it. After some time it became clear that the lines traced were not all in one line because of the Earth's rotation beneath the pendulum. If we ask what is beautiful about this experiment, there is an immediate answer: its visual features. We can regard the pendulum itself as beautiful, scientific equipment can certainly be beautiful: from chemical retorts and prisms to complicated instruments and detectors built in laboratories. Furthermore, the phenomena we study can also be beautiful: copper sulphate crystals, rainbows produced by prisms, and the microscopic structure of cells are all aesthetically pleasing to observe. But I argue that the ultimate beauty of experiments is found in their design and purpose. The beauty of Foucault's experiment goes much deeper beyond the visual, immediately accessible to us, features. It lies in showing the effects of the Earth's rotation, something important that hadn't been demonstrated before, in an ingenious, imaginative, and elegant way. The pendulum itself was beautiful, but the ultimate beauty of the experiment is a combination of its significance and its design.

To illustrate my argument I further analyse the Meselson-Stahl experiment designed to discover how DNA replicates. One of the reasons why this experiment is celebrated in science is because it is an example of a crucial experiment, it settled decisively the question on how DNA replicates, by selecting semi-conservative replication over the two alternative hypotheses that were entertained. A further aspect of the experiment's aesthetic value is not only what it taught us but *how* it did so and this later consideration concerns its design. Following the reasoning behind the experimental set up reveals the elegant design the experimenters created. The idea behind the experiment is considered beautiful and ingenious, the fact that by making the genetic material initially heavy and then light, Meselson and Stahl could extract and measure the weight of the genetic material though the next generations. It is in this idea that their design was original and elegant, they used the optimal materials and technique for the job. As such, the experiment integrates elegant design and involves innovative and creative thinking.

Next, I explore the asymmetry between experiments a century ago and experiments today, reflecting on the fact that while past experiments, like the Meselson-Stahl experiments, often involved a few scientists in a room, relatively cheap equipment and often the results could be perceived or established without lengthy interpretative work, today, experiments look rather different. I focus on the experiments ran at the Large Hadron Collider in CERN, which not too long ago detected the Higgs boson, vindicating the Standard Model. This experiment involves highly complex machinery and data analysis, it is a result of collaborative work between thousands of scientists, and the very boundary of the experiment transcends the borders of countries. Given their complexity and size, I ask whether these large-scale experiments fit with previous aesthetic ideals, or whether they can be praised for their aesthetic features and I argue that despite their complexity, large-scale experiments continue to be praised for their apt design and the creativity and originality they exhibit.

Last, I consider the case of some experiments that exhibit elegant and beautiful design and are well built for purpose, but obtain null results, asking whether they can be aesthetically appreciated. I examine the Michelson-

Morley experiment and argue that this experiment is aesthetically valuable even though it turned out that there is no such thing as ether. Contrary to Meselson and Stahl, who delivered an answer aligning to scientific expectation, the results of the Michaelson and Morley experiments were disruptive, but I argue that it is this disruptive nature of the result that was both aesthetically and epistemically valuable. It prompted the experience of wonder and disruption by identifying the limitations of our knowledge and prompted exploration of new ideas, leading to the development of Einstein's special theory of relativity and the abandonment of the Newtonian framework. The design was beautiful, the set up careful and original, the results were disruptive, surprising and awe-provoking. I propose that such experiments, just like many artworks that challenge our fundamental assumptions about ourselves and our place in the world can do, can deliver results that prompt us to reconsider our working assumptions. Their aesthetic significance is intricately related to our state of understanding and illustrates the diverse nature of the aesthetic experiences scientific products and artworks can elicit.

How to measure effect sizes for rational decision-making

Ina Jäntgen

Abstract

How to measure effect sizes for rational decision-making

Corresponding section: a (General Philosophy of Science)

Keywords: absolute measures, relative measures, effect sizes, trials, treatment choices

Short abstract

Absolute and relative outcome measures measure a treatment's effect size, purporting to inform treatment choices. I argue that absolute measures are at least as good as, if better than, relative ones for informing rational choices across choice scenarios. Specifically, this dominance of absolute measures holds for choices between a treatment and a control group treatment from a trial and for ones between treatments tested in different trials. This distinction has hitherto been neglected, just like the role of absolute and baseline risks in informing rational decision-making that my analysis reveals. Recognizing both aspects advances the discussion on reporting outcome measures.

Extended abstract

In empirical studies testing the effectiveness of treatments, the collected trial data is analyzed using outcome measures. These measures describe how the treatment and the outcome relate and are usually interpreted as measuring the effect size of the treatment. They provide information for policymakers, patients and others aiming to decide between treatments.

Not all outcome measures provide the same information though. In this talk, I focus on outcome measures for binary variables. Here, two classes of measures, absolute and relative ones, differ in how they describe a treatment's effect size. Consider the Heart Protection Study which tested the effectiveness of a cholesterol-lowering drug to prevent heart attacks and deaths (Heart Protection Study Collaborative Group, 2002). The study found a so-called relative risk reduction of 18 % of coronary death. The so-called risk difference was 1.2 %. Only the former effect size was reported, as is common in biomedical research. Yet, the difference in described effect size is striking. Aiming to decide on taking the drug, which effect size is informative for a decision-maker? The relative? The absolute? Or perhaps both? More generally, how should we measure effect sizes to inform rational decision-making?

In this talk, I argue that absolute measures are at least as good as, if not better than, relative ones for informing rational decisions across choice scenarios. More precisely, absolute but not relative measures provide the probabilistic information for choices between a treatment and the control group treatment from a single trial. For choices between treatments tested in different trials, we need information about the difference in the probabilities of the outcome of interest given the treatments, i.e. the difference in the absolute risks. Absent any knowledge about the probabilities of the outcome given control group treatments, the baseline risks, outcome measures do not provide this information. If we as deciding agents instead know the baseline risks, then the absolute risks can be calculated from both classes of outcome measures. If the baseline risks are known to

1

be equal across the trials but are themselves unknown, then absolute measures but not relative ones always provide sufficient information to choose. Overall, for informing rational decision-making, absolute measures dominate relative ones.

To establish these conclusions, I first introduce absolute and relative outcome measures. Then, I model two choice scenarios using decision theory, one involving outcome measures from a single trial and another involving outcome measures from distinct trials. Using these decision models, I identify an alternative to reporting outcome measures for informing decisions: reporting absolute and baseline risks. Still, reporting outcome measures is common practice in biomedical research, and, as I argue, could be warranted. Correspondingly, I analyze the conditions under which absolute or relative measures provide information for choosing treatments. As convincingly shown by Sprenger & Stegenga (2017), absolute measures but not relative ones always do so for choices between a treatment and a control group treatment. I argue that this argument does not hold for choices between treatments tested in distinct trials. Here, we need information about the difference in absolute risks to decide between treatments. To analyze when absolute or relative outcome measures provide this information, I distinguish between three epistemic situations, differing in how much we know about the baseline risks in the considered trials. I show that absolute measures are still at least as good as relative ones for informing choices across these epistemic situations. Overall, for informing rational decision-making, absolute measures dominate relative ones.

My analysis exposes the conditions under which both absolute and relative measures carry the probabilistic information a rational decision-maker needs, and when only absolute ones do so. Moreover, it identifies the role of absolute and baseline risks in rational choices. Recognizing both aspects advances the discussion on how to report effect sizes to inform treatment choices. In particular, Jacob Stegenga and his co-authors argue that only absolute measures but not relative ones are suited to inform rational decisions between treatments (Sprenger & Stegenga, 2017; Stegenga, 2018; Stegenga & Kenna, 2017). By contrast, I show when relative measures are just as good as absolute ones for this purpose. Still, I demonstrate that relative measures do not provide decision-relevant information that cannot be provided by absolute measures, including in choice scenarios Stegenga's work fails to consider. This finding strengthens the case against the need for relative measures, contra recent suggestions to report both absolute and relative measures (Hoefer & Krauss, 2021). Moreover, in biomedical research, most studies report only effect sizes measured in relative terms like the Heart Protection Study (Elliott et al., 2021). My results suggest that this practice could fail to inform treatment choices. Finally, I show that one could report absolute and baseline risks to guide decisions, providing an alternative to reporting outcome measures.

I conclude my talk with three principles for reporting outcome measures suggested by my analysis. These principles should be scrutinized in further work, in particular moving beyond the idealized perspective of decision-theoretic models.

Indicative bibliography

Elliott, M. H., Skydel, J. J., Dhruva, S. S., Ross, J. S., & Wallach, J. D. (2021). Characteristics and Reporting of Number Needed to Treat, Number Needed to Harm, and Absolute Risk Reduction in Controlled Clinical Trials, 2001-2019. *JAMA Internal Medicine*, *181*(2), 282–284.

Heart Protection Study Collaborative Group. (2002). MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: A randomised placebo-controlled trial. *Lancet*, *360*(9326), 7–22.

Hoefer, C., & Krauss, A. (2021). Measures of effectiveness in medical research: Reporting both absolute and relative measures. *Studies in History and Philosophy of Science Part A*, *88*, 280–283.

Sprenger, J., & Stegenga, J. (2017). Three Arguments for Absolute Outcome Measures. *Philosophy of Science*, *84*(5), 840–852.

Stegenga, J. (2015). Measuring effectiveness. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *54*, 62–71.

Stegenga, J. (2018). *Medical Nihilism*. Oxford University Press.

Stegenga, J., & Kenna, A. (2017). Absolute Measures of Effectiveness. In L. McClimans (Ed.), *Measurement in Medicine: Philosophical Essays on Assessment and Evaluation* (pp. 35–51). Rowman & Littlefield.

Distinguishing between selectable and circumstantial traits

Ciprian Jeler

Section: **Philosophy of Natural Sciences**

Title of the paper: **Distinguishing between selectable and circumstantial traits**

Keywords: philosophy of biology, natural selection, selectable traits, intrinsic propery, frequency-dependent selection

Short abstract:
There is surprisingly little philosophical work on conceptually spelling out the difference between the traits on which natural selection may be said to act (e.g. "having an above average running speed") and mere circumstantial traits (e.g. "happening to be in the path of a forest fire"). Here, I show that the two existing proposals as to how this distinction should be made are unconvincing because they rule out frequency-dependent selection. I then propose two new potential solutions, which share the idea that extrinsic properties of a particular type should be accepted as traits on which natural selection can act.

Extended abstract:

We intuitively accept that "having a high running speed" is a trait on which natural selection may be said to act (i.e. a selectable trait) while "happening to be in the path of a forest fire" is not. But there is surprisingly little philosophical work on conceptually spelling out this difference, i.e. on determining how we should distinguish selectable traits from merely circumstantial ones. I call this the "selectable traits problem" and, in this paper, I critically analyze the existing solutions to it and tentatively propose two solutions of my own.

First, I argue that the two existing solutions to this problem – proposed by Peter Godfrey-Smith and Pierrick Bourrat – are unsatisfactory. Godfrey-Smith's (2009) attempt to equate selectable traits with intrinsic properties of biological entities fails because it is too restrictive, ruling out, for example, cases of frequency-dependent selection. Indeed, in frequency-dependent selection, an individual of a given type does not have a fitness advantage or disadvantage because of the intrinsic property which allows us to pick out the types, but because of the frequency of its type in the population. It is not "having a particular (value for a) trait" that gives that individual a selective advantage or disadvantage; but "being rare (with respect that trait)" does. And this "rarity" is undoubtedly an extrinsic property, as it depends not only on the focal individual and its traits, but also on the traits of all the other members of the population. To claim that selection can only be said

to act on intrinsic properties is tantamount to claiming – against a long-standing consensus in evolutionary theory – that frequency-dependent selection is not natural selection. Moreover, I show that a particular attempt to reinterpret frequency-dependent selection in order to salvage Godfrey-Smith's solution is unconvincing because, in Sober's (1984) classical terms, it focuses on the trait on which there is "selection of" rather than on the trait that is "selected for."

On the other hand, Bourrat's (2015, 2017) solution, which consists in equating selectable traits with "intrinsic-invariable" properties, is even more restrictive and is thus unworkable. Moreover, I investigate whether Bourrat's caveat that even intrinsic-variable or extrinsic properties could be considered selectable traits if they are causally determined by intrinsic-invariable properties helps accommodate frequency-dependent selection. I argue that this is not the case.

Finally, I outline two possible solutions to the selectable traits problem that do not rule out frequency-dependent selection. These two potential solutions share the idea that extrinsic properties of a particular type should count as selectable traits. It is not my intention to decide which of the two solutions proposed here is preferable, but I argue that both of them are defendable and that each comes with its own advantages.

References

Bourrat P (2015) Distinguishing Natural Selection from Other Evolutionary Processes in the Evolution of Altruism. *Biological Theory* 10:311–321.

Bourrat P (2017) Explaining Drift from a Deterministic Setting. *Biological Theory* 12:27–38.

Godfrey-Smith P (2007) Conditions for Evolution by Natural Selection. *Journal of Philosophy* 104:489–516.

Godfrey-Smith P (2009) *Darwinian Populations and Natural Selection*. Oxford University Press, Oxford.

Sober E (1984) *The Nature of Selection. Evolutionary Theory in Philosophical Focus*. The University of Chicago Press, Chicago and London.

Quine's Semantic Holism: A Dispensable Theory?

Emerson Kang

Quine's Semantic Holism: A Dispensable Theory?

This paper will explore possible responses to one of the most successful objections to verificationism posed by W. V. Quine in the second part of "Two Dogmas of Empiricism" (1953,1960). Following Duhem, Quine denies that individual sentences have meaning. Quine's semantic holism entails that the meaning of individual sentences is context-dependent, and therefore that one sentence's meaning or truth is relative to its connection to all other sentences. It follows from this premise that individual sentences do not have verification conditions. In short, Quine demands that the verification of any sentence would require an infinite number of verification conditions of other sentences' meanings. On the other hand, because all of the sentence meanings are inter-connected, we would never know which conditions verify (or falsify) which sentences. In Quine's semantic holism, certain core beliefs, principles, and laws of nature—all of which can be expressed logically as sentences—are central to our knowledge of the world. These beliefs are not readily dispensed with, even when they are confronted by an experience that calls them into question. Meanwhile, other beliefs are more peripheral, and these peripheral beliefs are framed in terms of our more core beliefs. Could we save verificationism by undermining Quine's semantic holism? This paper will argue that Quine's semantic holism is, in fact, self-undermining. The arguments will draw on more recent treatments of Quine from Katz, Adler, and Chase. Quine's semantic holism will be shown to depend on a principle according to which any statement can be held true if other statements are modified accordingly—a principle which, however, would seem to apply its own negation. We can hold false the statement that "Any statement can be held true, if other statements are modified" if we are willing to modify our beliefs. For this reason, holism is subject to a self-referential paradox in terms of its central principle. The main argument of this paper will accordingly be that there is no way to defend semantic holism against the charge of relativism. Furthermore, the paper will argue that even if one could be philosophically comfortable with meaning-relativism, Quine's principle would be located very near the periphery of the web of our beliefs, not least because it seems tenuous in its self-referentiality. Meanwhile, the verificationist principle could be shown on empirical grounds to be much more near the center. That is, even accepting Quine's idea of core beliefs, most of us would appeal to a verificationist principle (as, for instance, when we empirically present something to clarify what a statement means) before we would appeal to a semantic holist theory of meaning. In fact, in the vast majority of everyday examples, some of which this paper will consider, it can be shown that verification itself is presupposed as a *de facto* theory of meaning. Thus, either verificationism stands on its own as a more coherent epistemological theory of meaning or, under Quine's holistic interpretation of meaning, verificationism would nevertheless serve as a more central criterion of meaning than Quine's principle.

Keywords:
1. W.V. Quine
2. Semantic holism
3. Verificationism

Michael Polanyi's tacit inference and socially engaged inquiry

Juozas Kasputis

General Philosophy of Science

Michael Polanyi's tacit inference and socially engaged inquiry

Key words: tacit inference, society of explorers, engaged inquiry

Many philosophically involved scholars have pointed out that M. Polanyi has been considered as 'outsider' within a tradition of philosophy of science. A former chemist turned into philosopher is not an extraordinary development in the history of science and philosophy. But M. Polanyi, as prominent natural scientist, has 'indwelled' (to use his famous term) epistemic issues from the genuinely unorthodox angle though with immense respect to the scientific tradition. After M. Polanyi's sociological/philosophical turn ('conversion' – another important term in his writings) both science and philosophy have received a fresh outlook on old debates. In general, the 1960s mark an intellectual upsurge in philosophy of science – M. Polanyi, Kuhn, Popper and others have responded to renewed concerns about scientific and social transformations. Though, M. Polanyi's ideas of personal knowledge and tacit inference have been overshadowed in the context of a famous philosophical debate between the followers of Kuhn and Popper/Lakatos regarding scientific revolution/change. 'Paradigm' and 'research programme' have become the established terms in philosophical vocabulary, but that still suffers from persisting issues of scientific formalism and specialism. In addition to this, scientific method of physics has been suggested as exemplary mode to "do science". It has created many challenges for philosophy of science in respect to social studies.

M. Polanyi's philosophical insights are still relevant for current discussions. His project of free society can be understood through a metaphor of the republic of science. In this sense, M. Polanyi has expanded the notion of scientific inquiry far beyond 'paradigm' or 'research programme'. It is an alternative approach to Popper's 'open society', which is haunted by controversial ideas of falsification and social engineering. There are sufficient reasons to employ M. Polanyi's ideas beyond epistemic framework – what suggests the concept of 'overlapping neighborhoods' between theory and practice, science and society. All in all, it makes the imagined society of explorers as potentially valid social/political program. M. Polanyi's criticism has addressed the reliance on formalism and detached mode of objectivity. His efforts to reintroduce social dimension into inferential schemes correspond with the idea of socially engaged inquiry. M. Polanyi has expressed a mistrust to the increasing preoccupation with automated chains of reasoning. In this regard, his discussion with Turing can have far reaching implications which were not fully recognized in the 1960s.

M. Polanyi has criticized the notion of formalized and detached knowledge as susceptible to the illusion of fixated perfectionism which leaves no space for human agency. Basically, his criticism attacks traditional formal logic exemplified by two modes of reasoning – induction and deduction. Inductive reasoning usually is defined as 'bottom-up logic' proceeding from the particulars to the universal. Accordingly, deductive reasoning is treated as 'top-down logic' proceeding from the universal to the particular. Following this thread of thought, inductive reasoning is merely the inversion of the deductive system thus sustaining a vicious mode of inquiry within limits of its own methodological comfort zone. M. Polanyi's logic of tacit inference diverges from explicit inferential schemes of induction and deduction. However, all modes of reasoning remain integrated in Polanyian framework of knowledge. M. Polanyi does not intend to oppose tacit and explicit knowledge, he rather criticizes the ideal of exact sciences which reduce tacit knowing to impersonal applications of theory. According to him, the logic must be viewed as embedded into social context. For the scientists and philosophers, it means a complete abandonment of isolated and detached position of observer.

Of course, M. Polanyi's ideas need to be critically evaluated as well. His view of society was strongly influenced by free market and traditionalist ideas of those days. Besides, there are some reservations about absolute freedom of scientific research – knowledge is power, thus unchecked power can cause much of trouble. Though, this also means that civic engagement of all citizens, including intellectuals, is crucial for the future of democracy. Dewey has warned back in the 1930's – too many people think of democracy as something given, which evolves in automatic fashion. We must remember it with regard to the current digital and technological transitions.

Epistemic Sustainability

Inkeri Koskinen and Samuli Reijula

Epistemic sustainability

Abstract for EENPS22

Philosophers of science have in the recent years increasingly paid attention to institutional structures and practices that surround and sustain scientific knowledge production. These structures and practices shape scientific communities and individual researchers' careers, and eventually, scientific knowledge. For various reasons – such as increasing the social accountability of science or ensuring social justice in scientific communities – philosophers have suggested that they should be altered. In this paper we offer a critical perspective to such suggestions from the viewpoint of the institutional epistemology of scientific knowledge. We argue that philosophers suggesting such changes should ensure that the changes are epistemically sustainable.

We start by introducing our notion of epistemic sustainability. Briefly, by epistemic epistemic sustainability we mean the ability of a knowledge producing system to continue producing reliable and relevant knowledge over time. Institutional structures can maintain or threaten the epistemic sustainability of a system.

We use the notion of sustainability in the context of the institutional epistemology of scientific knowledge. The field can be understood as an extension of social epistemology. Social epistemology of scientific knowledge focuses on the normative study of the social dimensions of scientific knowledge and practice. These practices do not, however, exist in an institutional vacuum. The institutional structures surrounding and sustaining science, both formal and informal rules of interaction (cf. Crawford & Ostrom 1995) – ranging from peer review and publishing practices to university governance and funding instruments – shape scientific knowledge production (Anderson 2006). Therefore the institutional epistemology of scientific knowledge focuses on the normative study of the institutional dimensions of scientific knowledge production.

In philosophy, the study of the social and institutional conditions of research has typically proceeded at a high level of abstraction and idealization. We believe that although high-level theorizing may serve some of its purposes, the institutional epistemology of scientific knowledge should adopt a less idealizing approach and approach its targets along the lines that are familiar to us from much of naturalistic philosophy of science and the philosophy of science in practice movement. To advance this aim, we introduce the notion of epistemic sustainability.

The notion of sustainability is most often used when talking about ecological or social issues, and sustainable development. The Brundtland Commission defined sustainable development in 1987 as development that "meets the needs of the present without compromising the ability of future generations to meet their own needs" (World

Commission on Environment and Development 1987). The temporal aspect of this definition is central also in a more abstract definition of sustainability: sustainability is the ability of a given phenomenon to endure, or the ability of a given system to persist across time (Colocousis et al. 2017).

By epistemic sustainability we mean the ability of a knowledge producing system to continue producing reliable and relevant knowledge over time. Scientific knowledge producing systems consist of elements such as research communities, formal and informal practices, institutional structures, accumulated knowledge and other intellectual, social, and material resources. The institutional structures that interest us in this paper are integral parts of the complex systems that produce scientific knowledge. Depending on how they are designed, they can contribute to maintaining the epistemic sustainability of the knowledge producing system, or threaten it.

A clarification is needed: not all problems arising from the institutional structures and practices surrounding and sustaining science are problems of epistemic sustainability. For instance, it may be possible to conceive an epistemically sustainable system that treats researchers unjustly. While in this paper we focus on epistemic sustainability, we do not think that a system like this would be desirable. We only argue that creating a fair and just system and creating an epistemically sustainable system are two distinct tasks, and succeeding in one does not automatically mean succeeding in the other.

Having introduced the notion, we use it for three purposes.

First, we use it as an analytic tool that captures a shared core in many apparently unrelated concerns expressed in many fields: the institutional structures surrounding current scientific knowledge production are causing epistemic trouble. This shared core between the various concerns is unmistakably normative. The concerns all include a demand that we call a demand for epistemically sustainable knowledge production. This analysis illustrates the kind of sustainability problems we can identify in contemporary science, and demonstrates the usefulness of the notion.

Secondly, we use the notion to present a critical viewpoint to the institutional countermeasures (e.g. establishing new practices in the allocation of funding, enforcing open data, changing department structures in universities, altering peer review practices, or demanding pre-publication of hypotheses) that have been suggested as solutions to the aforementioned problems. While we believe that many of these suggestions are valuable, we nevertheless think they should be critically scrutinised. The possibility that the suggested measures have unintended, negative epistemic consequences over time has not always been taken fully into account. We argue that an institutional remedy that fixes one sustainability problem but produces an entirely new one somewhere else is not satisfactory.

Finally, we extend our critical analysis to also other kinds of institutional reform ideas presented in the philosophy of science. Philosophers have suggested remedying various perceived problems in science by using institutional measures – problems ranging from wasteful practices in the allocation of research funding to a lack of social justice and accountability in science. These suggestions, when put into action, would introduce new institutional structures and practices in science, or change established ones. However, their interplay with other aspects of scientific knowledge production, and their unintended epistemic consequences over time, have not received enough critical attention. We argue that insofar as these suggestions are meant to be implementable, they should be based on a realistic understanding of the institutional context of scientific knowledge production as it is. It is not enough to demonstrate that a suggested institutional change, when put in practice, would remedy some problem. It is equally important to ensure that it will not lead to predictable outcomes that threaten the epistemic sustainability of scientific knowledge production.

Mapping Emotions in Scientific Experimental Practice

Anatolii Kozlov

Section: Philosophy of Natural Science

Title: Emotions in Scientific Experimental Research

Abstract

What is the place of emotions, if any, in scientific progress? Part of the difficulty related to this question is that emotions in science, although exist, are not very well documented. Emotions usually don't find their way into scientific reports or study books. Here, I conduct a structured sociological survey to capture the landscape of affective experiences associated with the practice of experimental research. I identify several inter-subjective factors that may shape the diversity of emotions nested within the experience of learning experimental results. I conclude by discussing epistemic implications of emotions in science and art.

keywords

Scientific practice

emotions

epistemology

survey

aesthetics of science

1

Is GPT-3 Language model a step towards Artificial General Intelligence ?

Roman Krzanowski and Pawel Polak

Philosophy of Cognitive and Behavioral Sciences
Is GPT-3 Language model a step towards Artificial General Intelligence ?

Keywords: Synthetic language models, cognitive systems, GPT-3, Artificial General Intelligence

GPT-3 (Generative Pre-trained Transformer 3) is an autoregressive language model that has produced human-like texts for a variety of domains. Some publications have suggested that GPT-3 displays a glimpse of artificial general intelligence (AGI). The study discusses a series of tests that demonstrate that statistical language models such as GPT-3 do not possess any form of intelligence one would expect from AGI system. A synthetic system with AGI capabilities should be able to answer complex questions from any domain, or at least to be able to reflectively evaluate its own response. In this study, we test GPT-3's capacity to produce short, coherent essays on philosophical themes. Our interest lies in whether GPT-3 can engage in some sort of philosophical analysis. Thus, we conducted a series of tests with the GPT-3 engine (January 20, 2022 edition) by presenting it with well-defined philosophical problems. We asked GPT-3 to respond to five relatively simple philosophical questions.[1] The questions were taken from the examination set for the University of London's (UofL) BA program (Examination papers and Examiners' reports, 2004). These were: Can we intelligibly claim that Sherlock Holmes does not exist? Is knowledge justified true belief? Must scientific explanations cite the cause? and Explain and evaluate the argument Descartes gives in the second meditation for the claim that he is a thinking thing.

GPT-3 model in the tested version was unable to create responses to questions that would indicate some understanding of texts. GPT-3 just produces responses without knowing what it is producing, so it cannot correct itself or improve its responses, because self-reflection is not a feature of the language model. GPT-3 produces a stream of words in a correct syntactic format, but it does not "know" whether an answer is good or bad. For GPT-3, the output is simply the output. The GIGO (garbage in, garbage out) rule does not apply because GPT-3 was certainly trained on a set of coherent texts, so producing irrelevant responses is a systemic feature. The causes of GPT-3's failings must be somewhere else other than the input feed. It may be that language cannot be abstracted from its cultural and social milieu (as it is what GPT-3 assumes) as suggested by some studies. In fact Chomsky model of language (universal grammar) assumes that language can be modeled as an abstract system (outside of its social and cultural grounding)

So, what do these results tell us the GPT-3 language model, and the language-study approach behind the GPT-model? We may speculate that linguistic tasks require more in-depth knowledge of the problem domain or a better understanding of the text beyond statistical correlations, even when they are derived from thousands of texts (as in the case of GPT-3). The GPT-3 language model remains a statistical synthetic model, nothing more. It was built under the assumption that given enough statistical data, the model will recover semantics, and given enough texts, the model will recover meaning. Based on the conducted tests, this assumption seems to be rather dubious. To be more cautious, however, we could say that this assumption was not confirmed in our tests. We may speculate that the considered problems require a deeper level of understanding, so they may be beyond this language model (or its current implementation) and

[1] By simple questions, we mean either that the questions are related to well-known philosophical ideas or that students at an early stage of philosophical training should be able to address them.

the GPT-3 approach to modelling language. The question we are left with is this: To produce a thinking artifact, which appears to be the ultimate goal of these linguistic models, will it suffice to absorb structural correlations, even on a superhuman scale (reading and digesting all of the sources that GPT-3 considered is certainly beyond any single person)? Or will this never be sufficient? In other words, would providing more training material and more computer power help? Throwing more resources at a problem is a sound military strategy, but it may not work with language. The results indicate that within the tested domain (i.e., philosophy), GPT-3 performed unpredictably, almost as if by chance. For some questions, GPT-3 produced answers that could be mistaken for a student's work, while for other questions, GPT-3 failed, with it simply generating a mixture of coherent and irrelevant, albeit grammatically correct, text. Any suggestions that GPT-3 may mark the dawn of a new era of synthetic philosophy, or the twilight of the traditional one, or a glimpse of AGI are simply not justified. We should not expect a flood of synthetic philosophy from GPT-3. Moreover, The tests with GPT-3 indicate that we should review the philosophical assumptions behind current AI systems as models of human cognitive functions.

Philosophy of Cognitive and Behavioral Sciences

Is GPT-3 Language model a step towards Artificial General Intelligence ?

Keywords: Synthetic language models, cognitive systems, GPT-3, Artificial General Intelligence

GPT-3 (Generative Pre-trained Transformer 3) is an autoregressive language model that has produced human-like texts for a variety of domains. Some publications have suggested that GPT-3 displays a glimpse of artificial general intelligence (AGI). The study discusses a series of tests that demonstrate that statistical language models such as GPT-3 do not possess any form of intelligence one would expect from AGI system. A synthetic system with AGI capabilities should be able to answer complex questions from any domain, or at least to be able to reflectively evaluate its own response. The tests with GPT-3 indicate that we should review the philosophical assumptions behind current AI systems as models of human cognitive functions.

Longino's Critical Contextual Empiricism and the feminist criticism of mainstream economics

Teemu Lari

e) History, Philosophy and Social Studies of Science

Longino's Critical Contextual Empiricism and the feminist criticism of mainstream economics

Keywords:

Helen Longino
critical contextual empiricism
science in society
values in science
dissent in economics

Extended abstract

Helen Longino's Critical Contextual Empiricism (CCE) is an influential normative account of the functioning of science (Longino, 1990, 2002). The account includes norms that epistemic communities such as scientific disciplines and their subfields should follow to support critical discussion that can uncover and challenge untenable background assumptions and value commitments. In this paper, I identify a tension in these norms and suggest a possible solution.

This is the tension: On the one hand, [A] the cognitive goals of epistemic communities should be open to criticism. In other words, the requirement of an ongoing process of "transformative criticism" applies not only to assumptions involved in research but also to questions about what kind of knowledge the community should aim to produce (Longino, 2002, pp. 164, 186). On the other hand, the contextualism of CCE entails that [B] appropriate criticisms of the commitments of an epistemic community must be "relevant to their cognitive and practical aims". Hence, a community need not be responsive to criticism that does not "affect the satisfaction of its goals" (Longino, 2002, p. 133). Clearly, a criticism to the effect that a particular goal should be abandoned or revised is anything but helpful in the pursuit of that very goal. Thus, the norm [B] seems to entail that the criticism of cognitive goals, which the norm [A] encourages, is necessarily irrelevant and requires no response.

The practical relevance of this tension is manifest in the criticisms that feminist economists have voiced against mainstream economics. In line with [A], some feminist economists argue that economics needs qualitative methods in the study of inequality and thus needs to count among its cognitive goals the pursuit of the kind of understanding provided by those methods (e.g. Doss, 2021; Figart, 1997). This criticism has evoked little by way of response from mainstream economists. However, such criticism is arguably not relevant for mainstream economists' pursuit of what they see as the goals of economics, so the neglect seems to be justified by [B] – a result certainly not intended by Longino.

I argue that the tension arises from conflating the "first-level" scientific discussion of an epistemic community with the "second-level" discussion about that epistemic community, its usefulness, and its role in the broader societal context. In the first-level scientific discussion, for scientific progress to take place, one must indeed take as given the overall goals and other core commitments of the epistemic community, as has been noted by Thomas Kuhn (1962) and others. The norm [B] is appropriate only at this level.

However, as [A] requires, there must be forums in which to discuss second-level questions like "what kind of knowledge should a certain discipline strive for?" The second-level discussion requires a revised norm [B*], according to which the relevance of criticism may be established by appealing to considerations like values or policy needs that may even conflict with the established goals of the community. The second-level discussion also needs to consider the institutional context such as the division of intellectual labor between epistemic communities and the way these communities exert influence in the society.

Thus, for example, when critics call for the incorporation of qualitative methods into economics curricula and admittance of qualitative research in top economics journals on the grounds that this is a way to develop a more thorough understanding of economic inequality, the mainstream economists opposing this view have a response duty. This is a second-level debate, so economists defending the status quo are not permitted to brush the criticism aside on the grounds that it is not relevant for what they see as the goals of economic research. In such a second-level debate, both parties of the debate should openly discuss the institutional context of the criticized epistemic community: how the division of intellectual labor among social scientific disciplines works, whether other social scientific disciplines (say, economic sociology) can better accommodate the qualitative study of inequality, and so on. The discussion also needs to consider that economics is a discipline with remarkable institutionalized power and prestige (Fourcade et al., 2015; Hirschman & Berman, 2014), so the produced knowledge may affect the society differently depending on which discipline it is produced by.

My analysis illuminates an issue that Longino has explicitly postponed for further study. She has wondered whether her "standards of argumentation" norm should be understood differently when the aims of research are debated, compared to when the debate concerns facts that the research tries to uncover (Longino, 2002, pp. 133, 212). The answer is: yes, there is a difference, and the distinction between first- and second-level discussion captures it. The strict requirement that all criticism invoke some standards of argumentation that are conducive to the community's goals entails that the room for acceptable criticism of the goals themselves is drastically reduced. This would be in stark contrast to the spirit of CCE – that the value commitments involved in scientific research should be uncovered and tried in an inclusive discussion.

References

Doss, C. (2021). Diffusion and Dilution: The Power and Perils of Integrating Feminist Perspectives Into Household Economics. *Feminist Economics*, *27*(3), 1–20. https://doi.org/10.1080/13545701.2021.1883701

Figart, D. M. (1997). Gender as more than a dummy variable: Feminist approaches to discrimination. *Review of Social Economy*, *55*(1), 1–32. https://doi.org/10.1080/00346769700000022

Fourcade, M., Ollion, E., & Algan, Y. (2015). The Superiority of Economists. *Journal of Economic Perspectives*, *29*(1), 89–114. https://doi.org/10.1257/jep.29.1.89

Hirschman, D., & Berman, E. P. (2014). Do Economists Make Policies? On the Political Effects of Economics. *Socio-Economic Review*, *12*(4), 779–811. https://doi.org/10.1093/ser/mwu017

Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago University Press.

Longino, H. E. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press.

Longino, H. E. (2002). *The Fate of Knowledge*. Princeton University Press.

Short abstract:

I identify and resolve a tension in Critical Contextual Empiricism (CCE) – the normative account of science developed by Helen Longino. CCE includes two seemingly conflicting requirements: The cognitive goals of epistemic communities may be criticized. But on the other hand, all criticism must respect the cognitive goals of that community to require a response. I demonstrate that the tension results from conflating the scientific discussion proper and the second-level discussion about science. Due to the conflation, CCE cannot adjudicate whether the unresponsiveness of mainstream economics to the criticism by feminist economists is permissible or not. My revision solves this problem.

Measurement in astrophysics: Can we be realists? And should we?

Anastasiia Lazutkina

Philosophy of Natural Science
Measurement in astrophysics: Can we be realists? And should we?
Astrophysics, measurement, realism, antirealism

Shorter abstract:

I argue that observational grounding of measurement in astronomy speaks in favor of a model-based view in measurement proposed by Eran Tal. An influential defense of antirealism in astrophysics has been given by Ian Hacking. I will show that astrophysics has progressed significantly enough in terms of measurement techniques since Hacking's paper that a new analysis of the issues is required. Sibylle Anderl has provided a partial response to Hacking that must be supplemented by considerations provided by Tal. My conclusion is that neither realism nor antirealism is vindicated by theses considerations.

Longer abstract:

In this work I will discuss how observations in astrophysics can be used to defend a realist or anti-realist position. Astrophysics is a branch of physics that combines theories from various branches of physics with astronomical observations to draw conclusions about the physical properties of astronomical objects. However, the measurement techniques significantly differ from those employed in other branches of physics due to the remoteness as well as the size of astronomical objects. We will consider arguments from Ian Hacking. Since Ian Hacking's criterion for realism is manipulability, he is led to endorse a modest antirealism about astrophysical phenomena. Anderl (2016) responds to Hacking by questioning whether the manipulability criterion for realism makes sense in general and whether there really are such fundamental differences between astrophysics and other natural sciences. Astrophysics is not just about passively gathering data from the cosmos but includes various aspects, such as deciding on the intended usage of the data prior to the observation, calibration of the telescope(s), calibration of data regarding atmospheric influence, etc. Anderl defends the realist position by appealing to the complexity of these activities, but this argument against Hacking's antirealism about astrophysics seems problematic. A further analysis of Anderl's other, more promising arguments is needed to decide on how successful her reply to Hacking is. What is clearly lacking in Anderl's approach, though, is discrimination between what we can know about different types of astronomical objects and about their properties. For instance, it seems we can know quite a lot about nearby stars and exoplanets by merely observing them, whereas peering inside the event horizon of a black hole, and thus confirming its existence as a singularity, is impossible. Between these extremes there are various shades of uncertainty about various types of objects and their properties. Realism about the existence and nature of nearby exoplanets seems no more problematic than realism about most scientific entities, but realism about the existence of singularities does not seem justified (perhaps general relativity breaks down in extreme conditions, but in a way that produces the observable event horizon). In conclusion, realism about astrophysics should perhaps not be a simple

yes/no-decision as Hacking or Anderl would have it, but a decision to be made about specific types of astronomical objects and specific types of properties, on a case-by-case basis. There is also a way to avoid both Hacking's antirealism as well as the realism Anderl seems to be pushing us toward: by considering some arguments for a perspectivist view of astrophysics and science in general, Hacking's view can be refuted without any kind of realist commitments. Specifically, this can be achieved by considering Eran Tal's account of scientific measurement. According to Tal, measurement is usually considered to provide especially strong evidence due to a special relationship it has to observation, and which scientific theorizing and modeling supposedly do not have. This view turns out to be mistaken, according to Tal's argument. Measurement surely does depend on observation in a special way, but this is not why it is considered to be superior evidence in comparison to theorizing and modeling, for measurement theory depends on both general scientific theorizing as well as idealizing data modeling. The relationship between measurement, observation, theorizing and modeling turns out to be reciprocal rather than one-way. But measurement does have a special kind of security as knowledge, which theorizing and modeling do not have. As Tal puts it: "Measurement outcomes are model-based predictors that have attained a high degree of security through various strategies, such as robustness tests." This claim must be qualified by what Tal says earlier: "[S]ecurity is not simply the inverse of uncertainty: a knowledge claim can be reported as having a low uncertainty but rest on assumptions that are likely to be revised in the foreseeable future. Such a claim would have a low security." In contrast, measurement outcomes as such are typically highly secure relative to the background theory and other models involved making refutable, unique predictions, but this does not mean the measurement outcomes in any way transcend the theoretical assumptions involved. The most secure measurement outcomes are thus those which only include assumptions that are unlikely to be revised in the near future. The current standard model of particle physics, as well as general relativity within the limits of a high-acceleration regime could serve as examples of such background theories. However, even such secure measurement outcomes remain theory dependent, and thus we have no more reason to trust those measurement outcomes as inherently more reliable which involve manipulation of the measurand compared to those in which measurement is more passive relative to the measurand. Accepting this conclusion would seem to involve rejecting Hacking's manipulability criterion of realism, but on the other hand, the acceptance of this conclusion does nothing to push us toward the type of realism that Anderl seems to want.

Can Laws of Nature be Categorical Properties?

Vassilis Livanios

Area: Philosophy of Natural Science
Title: Can Laws of Nature be Categorical Properties?
Keywords: DTA laws; Inference Problem; Categorical Monism; structural universals

The famous Inference Problem for DTA laws has a metaphysically loaded version (let us call this version MIP). If, for example, we focus on Armstrong's account of laws, the problem is how N(F,G) can metaphysically determine R(F,G), where N is the nomic necessitation relation and R is the relation between F and G that holds whenever the corresponding regularity (i.e., $\forall x, Fx \rightarrow Gx$) holds. In his "The Ultimate Argument Against Armstrong's Contingent Necessitation View of Laws" (2005), Bird brought to the fore another aspect of MIP. He argued that if we assume (as he thinks Armstrong has to assume) that the relation of nomic necessitation N is categorical, then no solution consistent with Armstrong's view about laws can be given to MIP. In brief, Bird argues that if Armstrong assumes that the state of affairs N(F, G) (which, according to him (1997), is also a first-order structural property) is ontologically *simple* and entails the corresponding regularity, then he must accept that one of the most important natural properties has a non-trivial inherent modal character. Bird thinks that simple categorical properties cannot have non-trivial inherent modal characters and so concludes that any version of categorical monism is inconsistent with N(F, G)'s determination role.

Bird's upshot can be refuted if N(F, G) is in fact a *complex* entity of which R(F, G) is a constituent. For all complex universals bear modal relations to their constituent universals irrespective of whether they are conceived as dispositional or categorical and so it seems that Armstrongian laws (conceived as complex first-order properties) can be categorical even though they entail regularities. Bird (2005, 152-3) anticipates this objection and argues against it. In particular, he points out that if R(F, G) is a constituent of N(F, G), then there must be an X such that N(F, G)=R(F, G)+X. Now X is either modally related to R(F, G) (i.e., entails by itself R(F, G)) or not. If the former, then effectively N(F, G)=X and we fail to explain what the constituents of N(F, G) are. Whereas following the latter option, we in fact hold the view that N(F, G) is a combination of R(F, G) plus 'something' independent of R, which, as Bird notes, is nothing other than a version of regularity theory of laws (i.e., a law is a regularity plus some additional factor) that Armstrong himself has convincingly argued against.

Bird's argument against the complex-N(F,G) possibility crucially relies upon the assumption that the only (relevant to the case under consideration) sense in which N(F,G) could be complex is to think the regularity fact R(F,G) itself as a constituent or a proper part of N(F, G). Given, however, Armstrong's mature view about laws (according to which laws are first-order *structural* properties that 'involve' causal connections at the property-level and are instantiated by causal sequences of particulars), this assumption is largely unmotivated. For within the Armstrongian context the most plausible explanation of N(F,G)'s alleged complexity is not that R(F, G) is a 'constituent' of it but rather that F, G and N are its 'constituents'.

If we embrace this interpretation of the complexity of Armstrongian law-structural universals, then the most salient question regarding Bird's critique is this: *can* there be modal relations (in particular, necessary connections) between the instances of a first-order categorical law-structural universal and the instances of the nomologically related properties which constitute the corresponding regularity? The answer, I suggest, is yes because the necessary connections between the instances of a structural universal and the instances of its constituents exist for all structural universals regardless of how we conceive them (that is, as categorical or dispositional). Therefore, unless one has reasons to think that law-structural universals are exceptions to this 'rule', it seems that there is nothing especially problematic concerning the above-mentioned necessary connections.

It might be objected that the above remark does not actually address Bird's objection because the latter says that, within a categorical monistic context, there is no metaphysical *explanation* of any necessary relation between a law-structural universal and its corresponding regularity. In response to that, I propose (on behalf of a proponent of DTA account of laws of nature) an essentialist explanation that is based on the following Essentialist View (EV) about laws:

(EV) A law-structural universal N(F, G) is essentially such that [necessarily, if N(F, G) is instantiated, then an instance of N(F, G) is identical to an instance of the regularity corresponding to R(F,G)].

In particular, I suggest that EV can be construed as the explanans of an (ultimate) explanation of the fact that, necessarily, if a law is instantiated, then an instance of that law is identical to an instance of its corresponding regularity. Since the postulated essence can be attributed to laws regardless of whether they are essentially dispositional or not, categorical monists who embrace a law-structural-universals kind of DTA view have the resources to deal with the explanatory problem. The conclusion is that Bird's challenge, even construed as a demand for a metaphysical explanation of the fact that raises MIP, can eventually be met.

Dreaming afterimages – subjective empiricism and self-experiments in the study of eidetic imagery and acousmatics in Central-European psychological tradition from Purkyně to Stanislav Vomela

Ivan Loginov

e) History, Philosophy and Social Studies of Science

Dreaming afterimages – subjective empiricism and self-experiments in the study of eidetic imagery and acousmatics in Central-European psychological tradition from Purkyně to Stanislav Vomela

keywords: psychophysiology, eidetic imagery, subjectivity, dreams, self-knowing

Extended abstract

This paper deals with the history of the research of eidetic imagery and dreams in psychophysiology of the 19th and early 20th centuries. It contributes to the history of eidetics as a field and aims to provide a better understanding of the context of psychological research on eidetic imagery and dreams in the 20th century. Another goal is to contribute to the study of the tradition of self-experiments in psychological research, here in relation to subjective empiricism and the concept of self-knowing of J. E. Purkyně.

The Czech scientist J. E. Purkyně is well-known for his research on subjective visual perception and afterimages, which he conducted at the beginning of the 19th century. In the early 20th century, Victor Urbantschitsch and E. R. Jaensch continued Purkyně's attempts to understand afterimages. Urbantschitsch was a German psychophysiologist who described the phenomenon of the subject not imagining the absent object but subjectively seeing it, which he called Anschauungsbilder (perceptual images). Jaensch, a German psychologist, named such phenomena eidetic and established eidetics as a new field or a research program. Eidetics then became quite popular in the first half of the 20th century but is hardly considered a stand-alone science or field today, although eidetic imagery itself is still of interest to contemporary psychologists.

Czech physician Stanislav Vomela inspired by Purkyně, Urbantschitsch, and Jaensch, studied his own eidetic experiences. His primary focus in this regard was on what he called the subjective perception of music – music heard but not imagined only by the subject. Later he chose the term acousmata to describe such phenomena and coined "acousmatics "as a name for the field of study. Directly inspired by Jaesch, Vomela understood acousmata as an auditory analogy of eidetic imagery. Methodologically he was inspired by Purkyně's subjective empiricism and his concept of self-knowing in the physiology of the senses.

Purkyně also tried to conduct subjective research of his dreams, though unsuccessfully. Stanislav Vomela followed in Purkyně's steps and studied his own dreams and eidetic and

acousmatic phenomena that he experienced while falling asleep, sleeping, and waking up. He called these hypnagogic hallucinations hypneidetic and hypnacousmatic phenomena. Later he came up with a peculiar theory explaining the nature of eidetic (and acousmatic) imagery (and sound). According to Vomela, hypneidetic phenomena were ontogenically primal, eidetic imagery thus being originally a part of the psychophysiological processes constituting sleep. He understood spontaneous and deliberate observation of such imagery by the subject in the state of wakefulness as a secondary effect of the existence of the original hypneidetic phenomena.

Self-experiments, the study of eidetic imagery, and the study of dreams were once connected – in Purkyně's research in the 18th century became reconnected by Stanislav Vomela and provided him with a unique set of tools to understand and explain the nature of eidetic imagery.

Short abstract

Czech physician Stanislav Vomela inspired by J. E. Purkyně, V. Urbantschitsch, and E. R. Jaensch, studied subjective auditory phenomena analogous to eidetic imagery, which he called acousmata. Following Purkyně's tradition, he also conducted subjective research on dreams, which led him to formulate his theory of the nature of eidetic phenomena. According to him, hypneidetic phenomena (eidetic phenomena perceived in dreams) were ontogenically primal, eidetic imagery thus being originally a part of the psychophysiological processes constituting sleep.

Batterman's minimal models: uniting global and local understanding.

Uzma Malik

Robert Batterman argues that his 'minimal' models often provide better understanding than more accurate detailed ones. I will explain minimal models using Batterman's own example in hydrodynamics (Batterman 2009/2021) to illustrate. I then develop an account of understanding— a 'local unification' account—that fits Batterman's claims, which I shall argue, serves global and local intuitions about understanding better than available alternatives.

According to Batterman's (2009) 'traditional view', the goal of mathematical modelling is a convergence between a model and reality: models are meant to present the most accurate and detailed mathematical representation possible of the phenomenon of interest. Idealisations are introduced only to be later de-idealised; one adds in details in order to de-idealise.

Contra the above, Batterman argues that a good model does not let details get in the way. The full details can "take something away" from a full understanding of the phenomenon of interest. In many cases, a certain kind of idealised model, which he calls a 'minimal model', is far better. Minimal models aim to expose 'common features' that systems with different detailed causal histories share. The idealising procedures followed to construct these models provide insight into why the different systems display these common features. A minimal model most "economically caricatures the essential physics".

Given that minimal models explain different kinds of cases with the same model, it is tempting to assimilate the understanding they provide into a unificationist understanding of the kind defended by Kitcher, Friedman etc. But doing so misses a large part of the attraction of minimal models. I argue that minimal models provide understanding by what I call 'local unification' which incorporates not only the global intuitions of conventional unificationism but local intuitions like those driving the causal mechanical view.

Understanding on the causal- mechanical view may be characterised as 'bottom-up' and local, this may be contrasted with the unificationist view characterised as 'top-down' and global. I argue that local unification understanding is different to standard unificationist understanding as well as the understanding provided by the causal-mechanical view. It is the way that idealisation is used that makes minimal models global; they are local because they start with particular phenomena and consider the local physics. I claim these two dimensions of understanding are entwined in local unification, inseparable and directly proportional to each other.

I close by contrasting my local unification with some similar seeming alternatives by Sorin Bangu (2017) and Michael Strevens (2004, 2008). Both see causality as the source of locality. Their accounts contain the same constituents as my local unification account but are organised differently. Bangu puts causality into a standard unificationist framework. Strevens' 'kairetic' account does the opposite. It incorporates unification within a causal framework. Though a charitable interpretation may see both accounts as hybrid, I will argue against this. I claim that my account is genuinely hybrid whereas Strevens' is more causal than unificationist and Bangu's is more unificationist than causal in accordance with the authors' intentions or context.

References

• Bangu, S. (2017). Scientific explanation and understanding: unificationism reconsidered. Euro Jnl Phil Sci 7: 103-126.

• Batterman, B. (2021). A middle way. Oxford University Press.

• ———————— (2009). Idealisation and modelling. Synthese 169: 427-446.

• ———————— (2002). A 'modern' (= Victorian?) attitude towards scientific understanding. Monist 83:2:228-257.

• Strevens, M. (2004). The causal and unification approaches to explanation unified – causally. Nous 38:1:154-176.

The challenges of constructing apt reference classes in biomedical research: on the example of racial categories

Joanna Karolina Malinowska

Section b) Philosophy of Natural Science

The challenges of constructing apt reference classes in biomedical research: on the example of racial categories

keywords: biomedical research, race, reference class, racial categories, medicine

Short abstract:

I reconstruct some of the methodological, epistemic, and ethical risks of using the category of race in biomedical research. I focus on the stage of constructing reference classes and how it influences the course of research. I discuss the issue of the US institutional guidelines for the use of racial categories and their impact on the construction of reference classes. I point out that the mere requirement to collect data on racial/ethnic groups affects research results and leads to the biologization of racial categories. Finally, I discuss some questions regarding the ethics of constructing reference classes in biomedical research.

Long abstract:

In principle, each person can be assigned to an infinite number of categories depending on the characteristic that will be its determinant (e.g., age, height, sex, gender, eye colour, favourite band etc.). In biomedical research, these categories are often called subgroups or reference classes. Constructing reference classes for research with human subjects ultimately relies on many evaluative decisions as well as on the historical and social context in which researchers are embedded (John, 2013; Ludwig, 2014, 2016; Reydon & Ereshefsky, 2022), although some suggest otherwise – that the choice of reference classes can be justified only by "natural" facts (Boorse, 1977; Veit, 2021) or epistemic values (Khalidi 2013).

Most scientists agree that many categories into which people can be divided (e.g., favourite actress or having bangs) are irrelevant to most biomedical research. Other, such as age, are universally recognised as significant. However, there are some categories the use of which is particularly controversial. One of them is the category of race.

In this paper, I reconstruct some of the methodological, epistemic, and ethical risks of using the category of race in biomedical research. I focus on the stage of constructing reference classes and how it influences the course of research. In particular, I discuss the issue of the US institutional guidelines for the use of racial categories and their impact on the construction of reference classes in biomedical research. I point out that the mere requirement to collect data on racial/ethnic groups affects research results and leads to the biologization of racial categories. For example, the current legal requirements in the US (in contrast to the UE) may enforce the assumption that racial categories are relevant because the institutions that regulate research require to use them. Researchers thus are encouraged to look for information to corroborate hypotheses about races/ethnicities and disease or

treatment options. Sometimes they may indeed find some correlations, if only because of the shortcomings of the methods they use, such as subgroup analysis (Sleight 2000; Lee 2009; Sun et al. 2012).

Finally, I formulate some questions regarding the ethics of constructing reference classes in biomedical research. Selecting and categorising participants for biomedical research not only co-shapes the research results (and their interpretations). It also may directly impact people's lives (John 2013). It is especially the case with such complex categories as race and ethnicity. These categories are not "discovered" by scientists. They are not objective "natural kinds" but instead are constructs used to obtain specific scientific, political, economic, or cultural goals. Therefore, I believe that it is required not only to transparently define what values are behind each scientific decision in this regard. I argue that it is necessary to formulate a framework for constructing social categories in biomedical research and practice. I also believe that such a framework must contain both: ethical and methodological guidelines, as they are inextricably linked in medicine. In this proposal it consists of the following elements: (1) scientific validity, (2) value transparency, (3) precise conceptualisation, (4) cautious application in medical practice and (5) potential social impact analysis.

Questions about all the above issues require reflection in the process of constructing reference classes in biomedical research. Let's return to using racial categories in science to clear this point. While analysing this problem, Dahlman (2017) asks whether the given reference class has a biological rationale and is homogenous (Dahlman 2017; cf. Greene 2019) (Ad. 1). The category of race (especially its official US legal interpretation) does not meet these conditions. However, eliminating it from science may make it impossible to analyse the health effects of racism. Therefore, most probably, some precise conceptualisation (Ad. 3) of racial categories that would help examine the above issue without perpetuating racial stereotypes or pseudoscience is needed. When it comes to social impact (Ad. 5), the main objection associated with using racial categories in biomedical research is the reification of old essentialist and hierarchical interpretations of human races. Stereotyping of patients resulting from its constant usage can significantly influence the behaviour of medical workers (Ad. 4). Moreover, as noted by Meissner (2021, p. 22), there are a lot of other potential harms in the use of folk racial categories in biomedical research, for instance, the creation and maintenance of dependency relationships, obscuring environmental racism or the creation and maintenance of trauma narratives that pathologize minorities. These issues contradict assumptions of fair selection and values (like social justice) associated with them. They can, however, be compatible with other non-epistemic values, such as economic ones (profit), e.g., when it comes to the pharmacological concerns investing in racial marketing (Salaz, 2010; Saha, 2015) (Ad. 2).

Boorse, C. (1977). Health as a theoretical concept. *Philosophy of science*, *44*(4), 542-573.

Dahlman, C. (2017). Unacceptable generalizations in arguments on legal evidence. *Argumentation*, *31*(1), 83-99.

Greene, C. (2019). Big Data and the Reference Class Problem. What Can We Legitimately Infer about Individuals?. *Computer Ethics-Philosophical Enquiry (CEPE) Proceedings*, *2019*(1), 7.

Khalidi, M. A. (2013). *Natural Categories and Human Kinds: Classification in the Natural and Social Sciences*. Cambridge University Press.

Ludwig, D. (2014). Disagreement in scientific ontologies. *Journal for General Philosophy of Science, 45*, 119–131.

Ludwig, D. (2016). Ontological choices and the value-free ideal. *Erkenntnis, 81*, 1253–1272.

Meissner, S. N. (2021). How does the consideration of Indigenous identities in the US complicate conversations about tracking folk racial categories in epidemiologic research?. *Synthese*, *198*(10), 2439-2462.

Saha, A. (2015). The marketing of race in cultural production. In *The Routledge companion to the cultural industries* (pp. 528-537). Routledge.

Sallaz, J. J. (2010). Talking race, marketing culture: The racial habitus in and out of apartheid. *Social Problems*, *57*(2), 294-314.

Sleight, P. (2000). Debate: Subgroup analyses in clinical trials: fun to look at-but don't believe them!. *Trials*, *1*(1), 1-3.

Sun, X., Briel, M., Busse, J. W., You, J. J., Akl, E. A., Mejza, F., ... & Guyatt, G. H. (2012). Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *Bmj*, *344*.

Veit, W. (2021). Biological normativity: a new hope for naturalism?. *Medicine, Health Care and Philosophy*, *24*(2), 291-301

Understanding selective semantic impairments

Andrei Marasoiu

c) Philosophy of Cognitive and Behavioral Sciences

Understanding selective semantic impairments

Which values does scientific understanding realize? Truth, explanatory power, predictive power, empirical adequacy, consistency, simplicity, perhaps others? I illustrate the discussion with a case study in cognitive neuropsychology, selective semantic impairments, arguing that researchers disbelieve the existence of selective semantic impairments. Standards for understanding these vary across disciplines (e.g., clinical cases vs. neural networks). And rational disagreements afford a limited plurality in which complexes of values are realized in understanding such impairments. All this undermines a purely ontic view of what is objective in understanding selective semantic impairments.

Keywords: scientific understanding; selective semantic impairments; ontic view of explanation; neural networks; clinical data

1. Introduction

Realism, contextualism and pluralism are often presented as mutually exclusive views; I will argue that this view is mistaken, and that scientific understanding realizes multiple epistemic values, differently in different contexts of research.

I apply the debate between realism, contextual and pluralism to scientific understanding for two reasons. First, understanding is itself puzzling, and crucial in the scientific enterprise by the lights of scientists themselves. Second, approaching scientific understanding doesn't beg the question against any of the views under consideration. If scientific knowledge was at stake, realism and contextualism might be favored over pluralism. If scientific beliefs or perspectives were at stake, contextualism or pluralism might be favored over realism.

My chosen case study in scientific understanding are selective semantic impairments. Their study mixes clinical, computational and large-scale functional models of semantic cognition. The recent history of this research project is interesting in its own right. I will argue it also calls into question whether pure forms of realism and pluralism can be supported, and favors instead a mixed view – one I will call "contextualist objectivism" – in which aspects of each are mixed to produce a metaphysics of science closer to the philosophy of scientific practice.

2. Selective semantic impairments

Consider patients who, for a variety of causes (physical injury, encephalitic herpes, stroke, senile dementia) come to lose their mastery of some concepts but not of others. For instance, patients who retain their ability to recognize, use and describe animate objects, but not so for inanimate objects – sometimes with the strange exception of fruits and vegetables (Farah 2004, p. 148). Other patients have the reverse problem: they are apt in dealing with artefacts, but can neither recognize nor understand, when they interact with living things, and with people, perhaps family, that they are doing so. These are extreme forms of what has sometimes been called "selective semantic impairments" or, interchangeably, "category deficits" (abbreviated "SSI's" henceforth).

I conjecture that three main stages could be distinguished in SSI study. They weren't recognized between the second half of the 19th and the first half of the 20th centuries, when research equivocally theorized semantic impairments, alexias, dyslexias and visual agnosias together.

A first stage in SSI research started with Elizabeth Warrington's 1975 article, approaching SSI's from a clinical standpoint. Warrington gave patients behavioral tests to see if they can recognize, name, draw various objects, and compared the results with the conditions patients suffered form.

A second stage in SSI study comes with Martha Farah and Jay McClelland's 1992 paper. Farah and McClelland use an artificial neural network to model human semantic memory. By targeted disruptions or destroying neural units, Farah and McClelland simulated patterns of impairment analogous to SSI patients. This was a proof of principle of how patients' semantic memory might work.

Following these developments, SSI literature grew significantly, relying on advances in the study of neural networks, especially network plasticity and the interaction between different networks, as well as increasingly flexible modeling of conceptual mastery (both at its lexical and its sensory ends). A third stage can be distinguished in SSI research, which I believe is well represented by the 2007 study by Lambon Ralph and Patterson. I believe we are still at this stage, which generalizes from SSI research by comparing alternative models of semantic memory in unimpaired patients.[1]

In brief, SSI research has seen generational progress. Put in terms of explanatory power, different explanatory levels are added at each stage: behavioral, clinical, computational, and psychological.

3. Truth and ontic explanations

I will develop a variety of contextualism: I hold that understanding SSI's realizes different epistemic values in different contexts. It follows that there is no unique epistemic value instantiated in all contexts of understanding SSI's, to which all other epistemic values boil down. In particular, I argue that truth is not a unique ultimate epistemic value at play in understanding.

An ontic account of scientific explanations (e.g. Strevens 2013) is extremely influential. Intuitively, unless what we explain genuinely exists, why bother explaining it? And what hold could we have via theories or models that might well be false? On the ontic[2] account of scientific explanation, untrue explanations don't genuinely explain.

I will now argue that the ontic account doesn't fit SSI's. According to Capitani et al.'s (2003) review, only 79 SSI cases have been documented between 1984 and 2001. The scarcity of

[1] In what concerns the development of neural networks, moving from the second to the third stage of SSI study is a placeholder for more subtle transformations in connectionist research. For an attempt at periodization of progresses in PDP modeling, cf. Clark's accessible (2001).

[2] I bracket Carl Craver's specific account of ontic explanation, on which what does the explaining are worldly mechanisms, quite independent of any thoughts or sentences that may or may not be true. I bracket Craver's important viewpoint because, if wordly entities and processes do play the roles assigned, then any description of their workings would have to be true. Even if such a description were never actually produced, it would still be there, in the offing, as it were. Undermining the tenability of any truth-based conception of explaining selective semantic impairments indirectly targets Craver's view as well. For an early discussion, cf. Machamer, Darden and Craver (2000).

extant empirical data is aggravated by their differing etiologies. Since my primary aim is to discuss selective semantic impairments, I can only conjecture that the situation generalizes. (Think of psychiatric disorders: their varying etiology may be bewildering.)

Moreover, Caramazza (1986) poses the problem of how to reconcile single-patient clinical case studies with statistically aggregated evidence coming from several patients. The problem is principled because each patient has a unique organism, and general systems' biology or neuroscientific models only apply to each organism partially, given their varying medical histories. This makes the inductive generalizations in SSI modeling trade statistic aggregation against fidelity to single-case studies in ways that cannot be accurately summarized by formal inductive rules.

Statistical generalizations might also equivocate over distinct kinds of cognitive and neural processing. Humphreys and Forde (2001) debate whether theorizing SSI's equivocates over disconnexion and degradation syndromes. It is one thing for patients to be unable to use the concept of a vase because they lost it completely, e.g. if degradation of a neural network no longer sustains high-level conceptual representations. It is quite another thing if patients cannot use the concept of a vase in visual recognition because that concept is inaccessible via bottom-up processing due to some problem en route with, say, shape-recognition or classification of vases as artifacts.

In sum, tentatively, the ontic view about the nature of scientific understanding doesn't do justice to these ongoing explanatory concerns.

References:

Capitani, E., Laiacona, M., Mahon, B. & Caramazza, A. (2003) What Are the Facts of Semantic Category-Specific Deficits? *Cognitive Neuropsychology* 20, pp. 213-261.

Caramazza, A. (1986) On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: The case for single-patient studies. *Brain and cognition* 5, pp. 41-66.

Clark, A. (2001) Connectionism. In: *Mindware* (pp. 62-83). New York: Oxford University Press.

Farah, M. & McClelland, J. (1991) A Computational Model of Semantic Memory Impairment: Modality Specificity and Emergent Category Specificity. *Journal of Experimental Psychology* 120, pp. 339-357.

Farah, M. (2004) *Visual Agnosia*. Cambridge, Mass.: MIT.

Humphreys, G. & Forde (2001) Hierarchies, Similarity, and Interactivity in Object Recognition. *Behavioral and Brain Sciences* 24, pp. 453-476.

Lambon Ralph, M.A. & Patterson, K. (2008) Generalization and Differentiation in Semantic Memory: Insights from Semantic Dementia. *Annals of the New York Academy of Sciences* 1124, pp. 69-76.

Machamer, P., Darden, L. & Craver, C.F. (2000) Thinking about Mechanisms. *Philosophy of Science* 67, pp. 1-25.

Strevens, M. (2013) No understanding without explanation. *Studies in History and Philosophy of Science* 44, pp. 510-515.

Warrington, E.K. (1975) The Selective Impairment of Semantic Memory. *Quarterly Journal of Experimental Psychology* 27, pp. 635-657.

Warrington, E.K. & Shallice, T. (1984) Category-Specific Semantic Impairments. *Brain* 107, pp. 829-854.

The Quest for Truth. Rethinking Scientific Understanding

Mariano Martín Villuendas

TITLE: The Quest for Truth. Rethinking Scientific Understanding

a) General Philosophy of Science

KEYWORDS: Scientific Representation, Idealization, Modeling, Veritism, Pragmatism

ABSTRACT:

Scientific understanding constitutes one of the major topics within the studies devoted to the philosophy of science. Nowadays, there is a general agreement in considering veritism as the only viable theoretical approach through which to address the characterization of this concept, thus being able to elucidate its role within the current scientific practice. The main aim of the communication is to assess the scope and validity of this theoretical approach. To this end, the fundamental theses on which this position is based will be presented and critically analyzed.

EXTENDED ABSTRACT:

Veritism constitutes a widely accepted position within the current landscape in the philosophy of science (Elgin 2017). This theoretical approach has been structured around three fundamental presuppositions. First, that truth constitutes both the ultimate goal of scientific practice and the fundamental criterion through which to assess the adequacy of epistemic products—e.g., theories, models, or explanations (Goldman 1999; Kvanvig 2003; Psillos 1999; Saatsi 2018). Second, that phenomena possess an ontic structure that is objective and thus independent of any pragmatic considerations (Craver 2007; Rice 2021; Strevens 2008). Third, that scientific knowledge, understood as justified true belief, is articulated through the grasping of explanations, which, to be adequate, must reflect the causal patterns or mechanisms that articulate the phenomenon of interest (Craver 2007; Kelp 2021; Khalifa 2017). To gain knowledge of a given phenomenon, cognitive agents must accurately represent the causal patterns or mechanisms of interest that account for the emergence of the phenomenon, at least those that make a difference—quasi-factivism (Bokulich 2016, p. 270; Kvanvig 2003). It is easy to note to what extent this view has exerted a profound influence on the main debates that have articulated—and still articulate—the contours of the philosophy of science: the realism-antirealism dispute (Psillos 1999; Saatsi 2018), the problem of scientific representation (Frigg & Nguyen 2020) or the problem of scientific understanding, the topic on which this communication will focus. This approach has been so predominant within the current philosophical landscape that even authors with a marked pragmatist tendency have assumed in their analyses the main theses underlying this position (Kitcher 2012; Potochnik 2017).

A central question is whether the adoption of this theoretical approach makes it possible to successfully address substantial philosophical problems such as those mentioned above. In order to answer this question, the present communication will take as its focus of analysis the problem of scientific understanding. Proponents of veritism have argued that the understanding of agents is adequate as long as it is factive. That is, in order to have genuine understanding, cognitive agents must grasp some kind of true information about the corresponding phenomenon—knowledge—which is structured through scientific explanations that, to be adequate, must reflect the corresponding ontic structure of the phenomenon, that is, its causal patterns or central mechanisms. The fundamental aim of the communication is to elucidate to what extent veritism provides a satisfactory answer to the question of what is scientific understanding, what is its value, as well as its place within scientific practice. Taking as a starting point the non-factive analyses conducted to date (De Regt 2017; Elgin 2017; Rouse 2015), the communication will show both the theoretical and practical inadequacy of veritism, calling into question the scope and validity of its analyses. To this end, I examine how this proposal is unable to accommodate the use of holistically distorted models—such as optimality models or simulation models—and to account for the existence of a plurality of alternative cognitive strategies and goals to causal or mechanistic explanation—such as mathematical explanations, probabilistic explanations or the exploration of spaces of possibility through the use of simulation models (Lange 2017; Reutlinger & Saatsi 2018).

The communication will be organized as follows. First, from an analytical point of view, I outline the fundamental characteristics of veritism, as well as its relation to the problem of scientific understanding. To this end, the complex theses that underlie it—factivism, representationalism, and explanationism—will be disentangled, and the main difficulties it faces will be presented. Second, I address the approach of the so-called "non-literalists" (Bokulich 2016; Frigg & Nguyen 2019; Rice 2021), a group of authors who have distinguished themselves by defending a more nuanced version of veritism in which they try to differentiate between representationalism and literalism, thus intending to overcome one of the main criticisms raised against this theoretical approach: the existence of holistically distorted models. According to these authors, it is possible to accurately represent a given phenomenon, albeit in a non-literal way. I explore to what extent this position fails to dissociate itself from the problems of standard literalist veritism. Third, two global arguments will be drawn against the ontological and epistemological presuppositions that underlie the veritistic approach to scientific understanding. I conclude by suggesting the need to abandon truth as the criterion to think about scientific understanding, adopting a non-factive and pragmatist approach instead.

REFERENCES

Bokulich, A. (2016). Fiction As a Vehicle for Truth: Moving Beyond the Ontic Conception. *The Monist,* 99, 260-279.

Craver, C. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neurosciences.* Oxford University Press.

De Regt, H. (2017). *Understanding Scientific Understanding*. Oxford University Press.

Elgin, C. (2017). *True Enough*. The MIT Press.

Frigg, R., & Nguyen, J. (2019). Mirrors without warnings. *Synthese*.

Frigg, R., & Nguyen, J. (2020). *Modelling Nature: An Opinated Introduction to Scientific Representation.* Springer

Goldman, A. (1999). *Knowledge in a Social World.* Oxford University Press.

Grimm, S. (2012). The value of understanding. *Philosophy Compass,* 7(2), 103–117.

Kelp, C. (2021). *Inquiry, Knowledge, and Understanding.* Oxford University Press.

Khalifa, K. (2017). *Understanding, Explanation, and Scientific Knowledge*. Cambridge University Press.

Kitcher, P. (2012). On the Explanatory Role of Correspondence Truth. In P. Kitcher (Ed.), *Preludes to Pragmatism: Towards a Reconstruction of Philosophy* (pp. 111-127). Oxford University Press.

Kvanvig J. (2003). *The Value of Knowledge and the Pursuit of Understanding*. Cambridge University Press.

Lange, M. (2017). *Because Without Cause. Non-Causal Explanations in Science and Mathematics.* Oxford University Press.

Potochnik, A. (2017). *Idealization and the Aims of Science*. The University of Chicago Press.

Psillos, S. (1999). *Scientific Realism. How Science Tracks Truth*. Routledge

Reutlinger, A. & Saatsi, J. (2018) *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations*. Oxford University Press.

Rice, C. (2021). *Leveraging Distortions. Explanation, Idealization and Universality in Science.* The MIT Press.

Rouse, J. (2015). *Articulating the World: Conceptual Understanding and the Scientific Image.* The University of Chicago Press.

Saatsi, J. (ed.) (2018). *The Routledge Handbook of Scientific Realism.* Routledge.

Strevens, M. (2008). *Depth: An Account of Scientific Explanation*. Harvard University Press.

The problem of causal inference from randomized trials

Mariusz Maziarz

The problem of causal inference from randomized trials

Short abstract:
The view that randomization does not balance confounding factors, which dates back to Worrall's (2002; 2007) papers is prevalent in philosophy of medicine. In contrast, medical researchers believe in the superiority of randomized studies over observational research. In my presentation, we revisit the role of randomization in causal inference and defend the view that randomization controls for the overall influence of confounding factors in the statistical sense. Additionally, we use a case study of a recent clinical trial to show that causal inferences in medicine rely on the assumption of randomization balancing confounders in the treatment and control groups.

Full abstract:
Worrall's arguments (2002; 2007) are prominent among philosophers of science but have not won the attention of medical researchers. We remain unconvinced and believe that Worrall's claims are at odds with probability calculus. In our paper, we argue that randomization balances the impact of confounders between the treatment and control groups in the statistical sense. That is, randomization asserts that the most likely division of participants into the treatment and control groups is such that the overall impact of confounders is equal and large imbalances unlikely. We support this claim in two ways. First, we use a toy example of toy-flipping exercise to show that randomization balances the influence of confounding factors in the statistical sense and explicate this sense of the claim. Furthermore, we analyze how the variability of the impact of confounders reduces with sample size and discuss some quantitative approaches to measuring this variability. Our analysis shows that the claims against the 'Millean balance' result from misunderstandings of how statistical inferences proceed: neither perfect balance of each confounder nor infinite samples are needed to estimate effect size confidence interval with high certitude. Second, we discuss the standard view on the role of randomization among medical researchers and use the example of the RECOVERY TRIAL to argue that the actual causal inferences in medical RCTs rely on the potential outcome approach that depends on the (probabilistic) notion of Millean balance. Third, we analyze the implications of randomization balancing the impact of confounders 'in the statistical sense' for causal inference and the trustworthiness of conclusions supported with RCTs.
The debate concerned with the question if randomization equalizes the influence of confounding factors on an outcome of interest, is at the heart of the debate about the role of randomization in causal inference. This discussion has started in response to the insufficiently examined view of the early proponents of evidence-based medicine that the results of randomized trials are superior to those reported by observational studies, at least for the assessment of treatment efficacy. Worrall (2002) notably argued that "[e]ven if there is only a small probability that an individual factor is unbalanced, given that there are indefinitely many possible confounding factors, then it would seem to follow that the probability that there is some factor on which the two groups are unbalanced (…) might for all anyone knows be high." (p. 324). His argument has been endorsed by several fellow philosophers. For instance, Thomson (2011) criticized the use of RCTs for causal inference repeating after Worrall that in actual clinical studies, treatment and control groups are heterogeneous for the reason that there are many confounders in medicine. He further added that "[I]n medicine, randomization is almost always gerrymandered (sampling is not from the

entire relevant population, some individuals assigned to a sample are removed after the fact, samples are adjusted to eliminate relevant differences observed after sampling or known to be likely from past experience (…). In addition, the assumption of homogeneity that is reasonably robust in Fisher's agricultural work is absent in medicine (…)." Borgerson (2009) argued against the privileged position of RCTs by pointing out that actual RCTs have only finite samples and may not reflect the average treatment effect of an ideal RCT with an infinite sample. Finally, Worrall (2007) himself further supported the criticism of randomization by arguing that "[t]here is no reason to think that [(…) average treatment effects (ATE) observed in individual studies agree with] the 'limiting average'" (p. 465) observable only if one re-randomized infinitely.

Others disagree with Worrall's objections to randomization. Cartwright (2010) distinguished between ideal RCTs that, by definition, assert the equal distribution of confounding factors between the treatment and control groups, and the actual studies. This argument was further extended by Deaton and Cartwright (2018), who criticized RCTs for their problems with extrapolation and analyzed the misunderstandings concerned with balance present in the literature but contended that randomization asserts the balance of confounders "in expectation". That is, if an RCT were repeated an infinite number of times, ATE would correspond to the true efficacy of an intervention. They also convincingly argued against Worrall's (2002; 2007) assumption that the equal distribution of each confounder is needed for sound inferences instead of balancing the average influence of confounders. Similarly, Philippi (2022) responded to the objection of Worrall (2002; 2007) by pointing out that estimating accurate ATE depends on the summary effect of confounders not differing significantly from zero instead of balancing each and every confounder. The strongest criticism comes from the statistician discussing seven myths of randomization. In his rebuttal of Myth 2, Senn (2013) differentiated between "a probability statement regarding the possible effects of possible imbalances (which is what with the usual statistical calculations provide) with a requirement for perfect balance (which does not exist)" and argued that any baseline imbalances do not undermine results due to the application of statistical testing. He further contended that Worrall (2002) confused a probability statement regarding the expected distribution of confounders with the requirement for perfect balance. He also rebuked Borgerson's (2009) claim by pointing out that the in-principle possibility of existing infinitely many confounders undermines the reason for randomization: even if there were an infinite number of confounders, it is their effect size that matters.

Still others agreed with Worrall's claim that balance in confounding factors is unattainable but defended randomization on other grounds. In Philosophy of Evidence-Based Medicine, Howick (2011) argued that randomization may not assert the equal balance of confounders due to limited sample sizes but they nevertheless are better at this task than non-randomized research designs and hence deserve being prioritized by the evidence hierarchies. La Caze et al. (2012) took issue with Worralls (2002; 2007) criticism of the epistemic virtues of randomization but they endorsed the purported falsity of the view that "random allocation controls for known and unknown confounders". Backmann (2017) pointed out that the potentially infinite number of confounding factors can be substantially limited on the basis of mechanistic evidence regarding treatment action. In his attempt at explaining the evidence-based medicine (EBM) evidence hierarchy, La Caze (2009) argued that randomization asserts higher internal validity in comparison to non-randomized interventional studies and observational epidemiology. Later, La Caze (2013) delivered a Bayesian justification for the higher trustworthiness of results stemming from randomized interventional studies.

References:

Backmann, M. (2017). What’s in a gold standard? In defence of randomised controlled trials. Medicine, Health Care and Philosophy, 20(4), 513-523.
Cartwright, N. (2010). What are randomised controlled trials good for?. Philosophical studies, 147(1), 59-70.
Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. Social Science & Medicine, 210, 2-21.
Fuller, J. (2020). The confounding question of confounding causes in randomized trials. The British Journal for the Philosophy of Science.
Fuller, J. (2021). The myth and fallacy of simple extrapolation in medicine. Synthese, 198(4), 2919-2939.
Howick and Mebius 2016 – review of the debate about the superiority of RCTs over observational studies.
Howick, J. H. (2011). The philosophy of evidence-based medicine. John Wiley & Sons.
Howick, J., & Mebius, A. (2016). Randomized trials and observational studies: the current philosophical controversy. In: Schramme, Th. And S. Edwards (eds.) Handbook of the Philosophy of Medicine. Pp. 873-886. Springer.
La Caze, A. (2009). Evidence-based medicine must be…. Journal of Medicine and Philosophy, 34(5), 509-527.
La Caze, A. (2013). Why randomized interventional studies. Journal of Medicine and Philosophy, 38(4), 352-368.
La Caze, A. (2013). Why randomized interventional studies. Journal of Medicine and Philosophy, 38(4), 352-368.
La Caze, A., Djulbegovic, B., & Senn, S. (2012). What does randomisation achieve?. BMJ Evidence-Based Medicine, 17(1), 1-2.
Philippi, C.L. (2022). There is no Cause to Randomize. Philosophy of Science. 89, 152-170.
Psillos, S. (2014). Causation and explanation. New York and London: Routledge.
Senn, S. (2013). A brief note regarding randomization. Perspectives in Biology and Medicine, 56(3), 452-453.
Senn, S. (2013). Seven myths of randomisation in clinical trials. Statistics in medicine, 32(9), 1439-1450.
Thompson, R. P. (2011). Causality, theories and medicine. Illari, Ph., F. Russo, and J. Williamon (eds.) Causality in the Sciences, 25-44. Oxford: Oxford University Press.
Worrall, J. (2002). What evidence in evidence-based medicine?. Philosophy of science, 69(S3), S316-S330.
Worrall, J. (2007). Why there's no cause to randomize. The British Journal for the Philosophy of Science, 58(3), 451-488.
Worrall, J. (2010). Do we need some large, simple randomized trials in medicine?. In EPSA philosophical issues in the sciences (pp. 289-301). Springer, Dordrecht.

Ante Rem Structuralism and the Semantics of Instantial Terms

Sofia Melendez Gutierrez

Abstract: *Ante Rem* Structuralism and the Semantics of Instantial Terms

Area: Formal Philosophy of Science and Philosophy of Mathematics

Keywords: philosophy of mathematics, mathematical structuralism, instantial terms, arbitrary reference.

Ante rem structures are universal entities exemplified by the members of equivalence classes under isomorphism. *Ante rem* structuralism is the theory mathematical entities are positions in *ante rem* structures (see e.g. Resnik [1997], and Shapiro [1997]). On this view, the natural numbers, for instance, are the positions in an *ante rem* structure exemplified by the finite von Neumann ordinals, the finite Zermelo ordinals, and all isomorphic systems. The rationale behind *ante rem* structuralism is the desire to endow mathematical singular terms with referents in the face of Benacerraf's contention that numbers cannot be identified with set-theoretical objects (1965).

Ante rem structuralists characterise *ante rem* positions as bearing only structural properties—viz. properties that can be exhaustively defined in terms of the relations that exist amongst the positions of a given structure. This characterisation of *ante rem* positions commits *ante rem* structuralists to the existence of entities that are indiscernible from others (see e.g. Burgess [1999, p. 288]; Keränen [2006, pp. 317-321]; and Shapiro [2012, pp. 380-381]): on their view, the two square roots of -1, for instance, are indiscernible, and so are all of the points in Euclidian space.

This result is significant for the *ante rem* structuralists for two reasons. First, it forces them to reject the principle of identity of indiscernibles on pain of absurdity: if the principle holds, there is only one square root of -1, and one single point in Euclidian space. Moreover, given that *ante rem* structuralism is meant to provide referents for the singular terms of mathematics, *ante rem* structuralists are strongly compelled to explain how 'i' refers to one of the square roots of -1 if they are indiscernible—and this is not at all an easy task (cf. Black [1952]). By definition, indiscernible entities are not located in spacetime and, hence, they are demonstratively inaccessible to us. Moreover, since they are qualitatively identical, every description satisfied by one of them is satisfied by both. Hence, neither can be singled out in order to establish it as the referent of any name.

Drawing from Roberts's work in linguistics on the semantics and pragmatics of definite noun phrases, Shapiro (2012) argued that the problem of reference to indiscernible entities may be solved by contending that the terms that refer to them are not names, but *instantial terms*. (2012, pp. 399-401). In my paper, however, I discuss several competing accounts of the semantics of terms of this kind, and argue that they are all untenable for the *ante rem* structuralist. First, I discuss Shapiro's own take on the matter. Briefly, on his view, $\ulcorner b \urcorner$ in $\ulcorner \Phi b \urcorner$ is empty until it is assigned a referent by means of a meta-theoretical function (2012, pp. 405-408). It is clear, though, that, for each one of the Φs, there is a function mapping $\ulcorner b \urcorner$ onto it. If the Φs are indiscernible, then these functions are indiscernible as well: they are all functions that can only be described as mapping $\ulcorner b \urcorner$ to one of the Φs; they cannot be distinguished by a specification of the particular Φ that they map $\ulcorner b \urcorner$ onto. All we can do, therefore, is to assert that a function f that assigns one of the Φs to $\ulcorner b \urcorner$ as a referent; but, if all such functions are indiscernible, then, by Shapiro's own lights, $\ulcorner f \urcorner$ cannot be a genuinely referring name—so it must be an instantial term.

According to Shapiro's own account of the semantics of instantial terms, $\ulcorner f \urcorner$ is empty until we assign a referent to it: we must assert that there is a higher-order meta-theoretical function g that assigns a referent to $\ulcorner f \urcorner$ from amongst the members of the set of indiscernible functions that we are considering—call it F. But, of course, for each one of the members of F, there is a higher-order function mapping $\ulcorner f \urcorner$ onto it; and, if the members of F cannot be discerned, then there is no way to specify which one of these higher-order functions is the one that assigns a referent to $\ulcorner f \urcorner$. If the members of F are indiscernible, then they cannot be individually denoted, nor referred to; but then the higher-order functions that map $\ulcorner f \urcorner$ onto each one of them are themselves indiscernible: they are all higher-order functions that can only be described as mapping $\ulcorner f \urcorner$ onto of the members of F; they cannot be distinguished by a specification of the particular member of F that they map $\ulcorner f \urcorner$ onto. All we can do, therefore, is to assert that a higher-order function g maps $\ulcorner f \urcorner$ onto one of the members of F. But, if all such members are indiscernible, then the singular term g cannot be a genuinely referring—and must, consequently, be itself an instantial term. On Shapiro's account, then, $\ulcorner g \urcorner$ is, by itself, empty; but we cannot tolerate that it remains this way, and thus we must assign it a referent.

This regress continues *ad infinitum*. Hence, the point where instantial terms are finally assigned referents is endlessly deferred, and never reached. Hence, they will remain empty forever. If they do not refer, then, trivially, they do not refer to any indiscernible entities. It follows that Shapiro's account of the semantics of instantial terms cannot help in solving the problem of reference to indiscernible entities.

In the second half of my paper, I briefly discuss four other theories about the semantics of instantial terms. Two of them (viz. Fine [1985]; and Breckenridge and Magidor [2012, pp. 377-378]), if true, turn futile the postulation of *ante rem* structures as the subject matter of mathematics. The *ante rem* structuralist, then, had better not endorse either of them. The others (viz. King [1991]; and Breckenridge and Magidor [2012, p. 384]) outright deny that instantial terms refer at all. If this is right, then they cannot help the *ante rem* structuralist solve the problem of reference to indiscernible entities; and, in particular, they cannot help them assign a referent to 'i' if the two square roots of -1 are cannot be discerned.

References

Benacerraf, P. (1965). What numbers could not be. *Philosophical Review* 74 (1):47-73.

Black, M. (1952). The identity of indiscernibles. *Mind* 61 (242):153-164.

Breckenridge, W., and Magidor, O. (2012). Arbitrary reference. *Philosophical Studies* 158 (3):377-400.

Burgess, J. P. (1999). Book Review: Stewart Shapiro. Philosophy of Mathematics: Structure and Ontology. *Notre Dame Journal of Formal Logic* 40 (2): 283-291.

Fine, K. (1985). Natural deduction and arbitrary objects. *Journal of Philosophical Logic* 14 (1):57-107.

Keränen, J. (2001). The Identity Problem for Realist Structuralism. *Philosophia Mathematica*9 (3): 308-330.

King, J. C. (1991). Instantial terms, anaphora and arbitrary objects. *Philosophical Studies* 61 (3):239-265.

Resnik, D. M. (1997). *Mathematics as a Science of Patterns*. Oxford University Press UK.

Roberts, C. (2003). Uniqueness in definite noun phrases. *Linguistics and Philosophy* 26 (3): 287-350.

Shapiro, S. (1997). *Philosophy of Mathematics: Structure and Ontology*. Oxford University Press.

Shapiro, S. (2012). An 'i' for an i: singular terms, uniqueness, and reference. *Review of Symbolic Logic* 5 (3):380-415.

The emergence of Earth System Science: paradigm shift or post-normal science?

Joao Mendes

SECTION A

The emergence of Earth System Science: paradigm shift or post-normal science?

Keywords: Earth System Science; Meta-methodology of Science; Thomas Kuhn; Imre Lakatos; Silvio Funtowicz; Jerome Ravetz

Abstract. At the turn of the 1980s to the 1990s, the so-called Earth System Science began to emerge, centered on the idea that «the Earth works as a single, complex and adaptive system, driven by the various interactions between energy, matter and organisms» (Steffen *et al.*, 2020). After early attempts in Forrester (1971) and Lovelock (1972), the contributions of Bretherton (1985), Earth System Sciences Committee-NASA Advisory Council (1986), and International Geosphere-Biosphere Program Global Changes (1986) were decisive to its establishment. This new epistemic framework has been increasingly accepted and consolidated in the last three decades by the scientific community. It appeared to replace the previously dominant epistemic framework of Environmental Science between the early 1960s and late 1980s. This was a framework that developed from growing public awareness and concern for environmental problems, mainly induced by the publication of books such as Rachel Carson`s *Silent Spring* (1962) and Paul R. Ehrlich`s *The Population Bomb* (1968), along with nuclear proliferation and growing concerns about the anthropogenic release of toxins and chemicals. Such framework has become an integrated, quantitative and interdisciplinary approach for the study of Earth`s environmental systems and solutions to their problems.
How to explain this change? Was it a scientific revolution, in Kuhn`s sense of a paradigm shift (Kuhn, 1996[3])? Or was it rather an epistemological transformation in the very way of doing science, particularly in the management of complex issues related to science, in the sense proposed by Funtowicz and Ravetz (Funtowicz & Ravetz, 1993)?
According to Sarkar (1980, p. 397), philosophers of science typically inquire at three levels: theories (hypotheses and conjectures) about the world (T); methodologies that evaluate these theories (M); meta-methodologies that have the important function of evaluating methodologies (MM). The issue I want to address here is at the MM level. It is therefore a question of the meta-methodology of science or, if you prefer, an exercise in the comparative philosophy of science. First, I will try to reconstruct in some detail the historical process of the emergence of Earth System Science. Next, I will analyze whether it is compatible with the Kuhnian conception of the succession of phases in the historical development of a science, that is, whether it corresponds to a scientific revolution and a paradigm shift. In a third moment, I will analyze whether it is compatible with Funtowicz and Ravetz`s conception of transition to a post-normal way of doing science in which uncertainty, axiological load and the plurality of legitimate perspectives are admittedly part of it. Finally, based on Lakatosian metamethodology, I will try to assess whether the process that led to the emergence of Earth System Science fits better in epistemological terms, that is, under which conditions we can know that one methodology is better than another (and not merely in logical terms, that is, under what conditions one methodology is better than another), with Kuhn`s methodological conception or with Funtowicz and Ravetz`s methodological conception (Lakatos, 1971).

References: Bretherton, F. (1985). Earth System Science and remote sensing. *Proceedings of the IEEE 73*: 1118–1127; Earth System Sciences Committee-NASA Advisory Council (1986). *Earth System Science: Overview: A Program for Global Change*. Washington, DC: The National Academies Press; Forrester, J. (1971). *World Dynamics*. Pegasus Communications; Funtowicz S. & Ravetz, J. (1993). Science for the Post-Normal Age. *Futures, 25*, pp. 735-755; International Geosphere-Biosphere Program Global Changes. (1986). *The international geosphere-biosphere programme : a study of global change : final report of the Ad Hoc Planning Group ICSU 21st General Assembly, Berne, Switzerland 14-19 September, 1986*. Stockholm: International Geosphere-Biosphere Programme: A Study of Global Change (IGBP); Kuhn, T. (1996). *The Structure of Scientific Revolutions*. 3rd ed. University of Chicago Press; Lakatos, I. (1971). History of Science and Its Rational Reconstruction. In R. Buck & R. Cohen (eds), *PSA 1970. Boston Studies in the Philosophy of Science, vol 8* (pp. 91-136). Springer; Lovelock J. (1972). Gaia as Seen Through the Atmosphere. *Atmospheric Environment, 6*, pp. 579-580; Sarkar, H. (1980). Imre Lakatos' Meta-Methodology: An Appraisal. Philosophy of the Social Sciences, 10(4), 397–416; Steffen, W., Richardson, K., Rockström, J., Schellnhuber, H., Dube, O., Dutreuil, S., Lenton, T. & Lubchenco, J. (2020). The emergence and evolution of Earth System Science. *Nature Reviews Earth & Environment 1*, pp. 54-63.

Questioning Rein Vihalemm's model of phi-science

Ave Mets

Section: a) General Philosophy of Science

Title: Questioning Rein Vihalemm's model of phi-science

Keywords: phi-science; Rein Vihalemm; scientific practice; epistemic pluralism

Rein Vihalemm conceptualised scientific methodologies as physics-like or phi-sciences and as non-phi-sciences (here meant in the narrower sense as natural sciences). An essential characteristics of phi-sciences is their a-prioristic approach to their research object, where theoretical-mathematical idealisation is constructed as the formalisation of the science's laws, and the material world to be accounted with it is constructed as laboratory experiments or extra-lab apparatus, thus being subsumed to the reign of those a priori mathematical tools, or adapted to theory. Non-phi-sciences or natural sciences, to the contrary, have to attend to their research object, its details and idiosyncrasies—their aimed material world being the open living world constantly adjusting to "external" influences, and adapt their theory to them—a reverse approach compared to the phi-sciences. This characterisation helps, inter alia, explain why biological sciences have hard time formulating predictive laws of nature of the type that physics does.

Although Vihalemm (2016) mentions, not disapprovingly, the idea (referring to Toulmin) that most sciences are 'a mixture of natural history and physics' (being both descriptive and explanatory, correspondingly, to certain degrees), he does not go into details of this mixture even in his pet case of chemistry. However, taking a closer look at the diversity of cognitive activities in sciences with the aim of determining their character as either phi- or non-phi-scientific, and hence of determining their host science itself as one or the other, one gets into trouble: the conceptual division is narrow, unspecific, and impoverished. The impoverishment means that some types of science, such as productive, that chemistry itself primarily is, are left out. Here I focus on the narrowness and unspecificity.

By 'unspecific' I mean that the subject to the adjective 'phi-scientific' cannot be clearly made out. Vihalemm writes about sciences (disciplines) as this subject, but as already mentioned, there are epistemic activities in sciences whose spectrum is much richer than construction or description, e.g. different kinds of computer simulations, statistical methods, evaluation of measurement errors and uncertainties, material extrapolations (i.e. not for theoretical hypothesis but explanation of material findings, like fossils explained on the basis of current organisms; see Currie 2018), etc.. Due to the neglect of this richness, the division into phi- and non-phi-sciences appears narrow, seemingly clearly applying to only theoretical mechanics and some fields of biology. But even the practices and activities themselves, when recognised as variously layered as Hasok Chang does (2012), have aspects which could be described as phi- or non-phi-scientific. E.g. a quantitative measurement—a phi-scientific activity—has as a sub-operation determining possible sources of error—a non-phi-scientific activity. Thus Hanne Andersen (2016) considers cognitive convergence and divergence of research operations and activities within research fields, but it seems, again, that what she regards are interdisciplines, obtaining their input from different disciplines with different domains (e.g. geomicrobiology, integrative neuroscience), leaving open the option that each of those disciplines involved is still methodologically fairly uniform.

Possibly Vihalemm's methodological division should primarily be assigned to the outcomes of research activities—the theories and laws or regularities accumulated in a particular field, and not activities themselves. This, however, is an unlikely interpretation, firstly for his advocated practical realist approach to science, secondly for his support for Ronald Giere's contention that all sciences build models, and models always only capture some aspects of the modelled phaenomenon, which contradicts the non-phi-scientific pretension of attention to detail and idiosyncrasies. In addition, the outcomes importantly depend on the activities themselves and which aspects of them are valued. Nancy Tuana (1996/2014) gives telling examples of how carrying out scientific activities in one way or another, due to the scientist's background beliefs and attitudes, influences the product of the scientific practice in dramatic ways. Among those influences are the extent to which a discipline turns out as phi- or non-scientific, notoriously that above mentioned, exemplary non-phi-science biology.

I conclude that the model of phi- and non-phi-science needs revaluation as to its limits and application domain in order to be a useful model in philosophy of science.

Bibliography

Andersen, Hanne. 2016. Collaboration, interdisciplinarity, and the epistemology of contemporary science. *Studies in History and Philosophy of Science* 56, pp. 1-10

Chang, Hasok. 2012. *Is Water H2O? Evidence, Realism and Pluralism*. Springer

Currie, Adrian. 2018. *Rock, Bone, Ruin. An Optimist's Guide to the Historical Sciences*, MIT Press

Tuana, Nancy. 1996/2014. Revaluing science: starting from the practices of women. In Scharff, R. and Dusek, V. (eds.), *Philosophy of Technology. The Technological Condition: An Anthology*, Wiley Blackwell, pp. 161-170

Vihalemm, Rein. 2016. Science, ϕ-Science, and the Dual Character of Chemistry. In Eric scerri and Grant Fisher (eds.), *Essays in the Philosophy of Chemistry*, Oxford University Press, pp. 352-379

The Shaping of Venn Diagrams

Amirouche Moktefi

Section: f) Formal Philosophy of Science and Philosophy of Mathematics

And/or section e) History, Philosophy and Social Studies of Science

The Shaping of Venn Diagrams

Keywords: Venn diagram, Euler diagram, Boolean logic, History of logic

Short abstract (100 words): Venn diagrams are widely used. They first appeared in 1880 to tackle elimination problems in logic. Little is known on their process of invention. Venn presented them as an "almost entirely new device" and said he has "never seen any hint at such a scheme". We argue that several such hints existed but Venn often overlooked them or failed to acknowledge their merits. We discuss Eulerian conventions that express uncertainty regarding the knowledge conveyed by propositions. Then, we expose Boolean step-by-step methods for the representation of propositions. Venn's merit was to combine these two traditions to achieve the desired scheme.

Extended Abstract (633 words): Venn diagrams are widely used in modern scientific literature. It is well known that John Venn first published these celebrated diagrams in 1880. They were designed for the solution of elimination problems, which were essential to the new Boolean logic. Boole famously expressed propositions in the form of equations, and hence, reduced logical problems to systems of equations. Elimination simply consists in determining what conclusion follows from any set of propositions offered as premises, involving any number of terms. Boolean logicians designed symbolic, diagrammatic and even mechanical solutions to this problem. They also often engaged in a friendly, and sometimes not-so-friendly, contest by comparing their notations and methods. This 'research program' required new types of diagrams since the older schemes, notably Euler diagrams, were unsuitable for the purpose. Venn championed Boole's logic. His diagrams precisely tackled elimination and opened the way to a family of 'Boolean' diagrams intended to address such problems when the number of terms increased.

Venn's method of representation consisted in a two-step procedure. First, one draws a primary diagram that represents the combinations between the terms involved in a proposition or an argument. Then, one adds syntactic signs on this general framework to indicate the state of the compartments. For instance, shading a compartment expresses its emptiness. Venn says

very little on the journey that led him to these diagrams. All we are told is that he "tried at first, as others have done, to represent the complicated propositions [of Boole] by the old [Eulerian] plan; but the representations failed altogether to answer the desired purpose; and after some consideration [he] hit upon the plan here described" (J. Venn, 'On the diagrammatic and mechanical representation of propositions and reasonings', *Philosophical Magazine* 10, 1880, p. 4). Although this narrative suggests a sudden or lucky insight at work, it also reveals the two ideas that Venn was investigating when he hit upon his diagrams, namely Eulerian diagrams and Boolean logic.

The aim of our talk is to determine what Venn diagrams owe to these two traditions. On the one hand, Venn's diagrams appear to be an amended version of Euler's. As early as 1881, William Stanley Jevons described Venn's scheme as "a complete and consistent system of diagrammatic reasoning, which carries the Eulerian idea to perfection" (W. S. Jevons, 'Review of John Venn's *Symbolic Logic*', *Nature* 24 (611), 1881, p. 233). Yet, Venn regarded his diagrams as a "special, and almost entirely new device" (L. M. Verburgt (ed.), *John Venn: Unpublished Writings and Selected Correspondence,* Springer, 2022, p. 70). On the other hand, Venn acknowledged that his method was "founded" on Boole's system, but insisted that it was not "in any way directly derived from" Boole himself. Indeed, the latter "does not make employment of diagrams himself, nor does he give any suggestions of their introduction" (J. Venn, *Symbolic Logic*, Macmillan, 1881, p. 104).

Venn explained that he was searching for a "new scheme of diagrammatic representation which shall be competent to indicate imperfect knowledge on our part; for this will at once enable us to appeal to it step by step in the process of working out our conclusion" (Venn 1880, *ibid*, p. 4). Venn's diagrams precisely fulfilled this purpose. Venn insisted on their originality and stated that he has "never seen any hint at such a scheme" (*ibid*, p. 4). In this talk, we identify several hints that opened the way to Venn's invention, even if the latter often overlooked them or failed to acknowledge their merits. We first explore post-Euler conventions that expressed uncertainty regarding the knowledge conveyed by traditional propositions. Then, we discuss how Boolean logicians developed a step-by-step method for the representation of propositions and sets of propositions. Then, we argue that Venn's merit was precisely to combine these two traditions to achieve the desired scheme.

Mental kinds and practical realism

Bruno Mölder

c) Philosophy of Cognitive and Behavioral Sciences

Title: **Mental kinds and practical realism**

Keywords: mental entities, practical realism, interpretivism, natural kinds, self-identifying objects,

Is there a consistent position on which mental entities do not wear their labels on their sleeves, whereas some natural entities do?

A non-global interpretivist conception could be viewed as such a position. In this view, mental states do not wear their labels on their sleeves, that is, they do not bear their mental specification intrinsically. On the contrary, meriting a mental specification is an extrinsic state of affairs – this depends on various extra-mental factors, and interpretation is required to pick out the suitable mental specification. In this interpretivist conception, the possession of mental properties is recognition-dependent, whereas the recognition-dependence does not hold across the board for all properties. Whether an object has certain natural properties is not constitutively dependent on interpretation.

The presentation proceeds from this starting point and answers the criticism made of such a position by Estonian philosopher of science Rein Vihalemm from the point of view of practical realism. In Vihalemm's view, the mental is not special in this respect, for there are no entities that wear their labels on their sleeves.

I outline some views that influenced Vihalemm, such as Hilary Putnam's internal realism and Ilkka Niiniluoto's critical scientific realism. On the backdrop of Vihalemm's criticism of entities wearing their labels of their sleeves is Putnam's rejection of the existence of "Self-Identifying Objects", i.e., objects that are individuatively-independent from our conceptual schemes. Niiniluoto also rejects the self-identifying objects but allows the talk about the mind-independent WORLD, which can be categorized in various ways. Vihalemm, in his practical realism, proceeds from Niiniluoto, but stresses the role of scientific practice and allows access to THE WORLD only through the theories constructed by us. He also takes natural kinds to be relative to the models we have constructed.

In this background, I discuss two options concerning the special status of mental entities.

1. Accept practical realism. Even though there are no self-identifying objects and natural kinds are also dependent on our individuation practices, mental kinds would still be even more interpretation-dependent than natural kinds.

2. Reject practical realism, in particular the view that the individuation of natural kinds depends on our theories and scientific practices. The world would divide up into natural kinds even if there had been no scientific practice, but there would be no mental kinds without the practice of folk psychology.

In my talk, I argue for the second option, by outlining the role of folk psychological framework and practice in making up the mental classifications and providing reasons why mental kinds should not be regarded as natural kinds.

Construct Validation and Pluralism in Psychiatry

Daniel Montero Espinoza

Title: Construct Validation and Pluralism in Psychiatry

1. Problems with the DSM Categories

Unlike in most other branches of medicine, in psychiatry, clinical instances of a (mental) disorder are identified almost entirely on the basis of symptomatology. The search for the biological bases of mental disorders –on the other hand—has been largely unsuccessful until nowadays. The syndromal (i.e., symptom-based) definitions of psychiatric conditions found in the *Diagnostic and Statistical Manual of Mental Disorder* (DSM) have become discredited due to two pressing clinical phenomena. First, patients often meet the criteria for more than one –sometimes several— psychiatric diagnose, this is known as *comorbidity*. This phenomenon raises the worry that perhaps too much emphasis has been placed on studying specific psychiatric disorders in isolation from other neighboring psychiatric conditions with which they might share similar etiologies. Second, because psychiatric diagnoses are defined as sets of symptoms and patients can meet the diagnostic criteria in several different ways, it is often the case that two patients share the same psychiatric diagnosis despite having almost no symptoms in common, this is known as *heterogeneity*. For instance, to meet the diagnostic criteria for major depression disorder, a person must have at least five out of nine symptoms for at least two weeks. Thus, two patients might be diagnosed with major depression while having only one symptom in common. Such degree of heterogeneity might prevent researchers from identifying some of the distinctive biomarkers of different psychopathologies. It is widely held that, given their high degree of heterogeneity and comorbidity, current psychiatric categories have low *diagnostic validity*, i.e., the set of diagnostic criteria comprised by each psychiatric category does not lead to proper identification of clinical instances of a psychiatric construct.

2. The Research Domain Criteria (RDoC) Initiative

One recent attempt to overcome the current crisis in psychiatric research is the Research Domain Criteria (RDoC) launched in 2010 by the National Institute of Mental Health –currently in development. RDoC is a research framework –organized as a matrix—that seeks to investigate mental disorders by collecting data from several levels of analysis that include genomics, molecules, neurocircuitry, behavior, and self-reports. Such data informs basic *dimensions* of functioning that cover a wide range of human behavior –-spanning from normal to abnormal behavior. Each dimension of functioning comprises different constructs which, in turn, represent different *aspects* of the overall range of functioning defined by the dimension. For instance, the dimension "negative valence system" –which encompasses the range of responses to aversive situations—comprises the constructs: acute threat, potential threat, sustained threat, loss, and frustrative nonreward. Each construct is intended to be measured –if methods are available—at the different levels of analysis mentioned above. While remaining agnostic about the DSM categories (Cuthbert, 2014), the RDoC's matrix introduces a new ontological framework constituted by constructs that describe cognitive functions.

One of the main challenges that RDoC faces is validating its constructs at the several levels of analysis of the matrix. According to Cronbach and Meehl (1955), validating a construct amounts to show that the construct fits in a nomological network. In their view, to "make clear what something is" –scientifically speaking—means to set forth the laws in which in occurs. "Learning more about" a theoretical construct is a matter of elaborating the nomological network in which it occurs, or of increasing the definiteness of the components (ibid). However, as Sullivan shows (2009; 2010), there is often little agreement among researchers upon the conceptualization of a cognitive phenomenon (construct) that they intend to measure. Moreover, in some cases, the measuring methods differ enough to rise questions regarding the extent to which they measure the same construct. This kind of disagreement poses a challenge for RDoC when it comes to validate its constructs across the multiple levels of its matrix.

3. Construct Validation and Methodological Pluralism

Several authors (Aftab & Jerotic, 2021; Bueter, 2019; Vintiadis, 2015) have claim that scientific pluralism –most prominently, methodological pluralism—is the only way forward in psychiatry. According to this view, "no single scientific method is satisfactory, and we require a pluralism of empirically rigorous methods and perspectives. Methodological pluralism requires the recognition of strengths and limitations of any given method, and the questions for which it can provide the best answers" (Ghaemi, 2007 as in Aftab & Jeroic, 2021, p. 538).

In my paper, I will focus in RDoC's framework to argue that there is a tradeoff between methodological pluralism and construct validation in psychiatric research, namely, the more diverse the methods employed, the harder it becomes to achieve construct validation. In this context, I seek to answer to the question *how much pluralism in psychiatric research?* On the one hand, we want to remain as pluralist as possible, while on the other hand, we want our constructs to be as valid as possible across multiple levels.

In answering to this question, I maintain that pluralism is beneficial insofar as (1) there is enough agreement about the conceptualization of the constructs being used across different levels of analysis and how they are measured; and (2) constructs are validated by means of (increasingly) specifying their place in a nomological network. Regarding (1), I claim that agreement upon the conceptualization of constructs need not be absolute, but wide enough to enable future refinement. In support of (2), I argue that, in psychiatry, it is not enough to be able to predict – for instance—the course of some psychiatric condition, but we need to understand the kinds of interactions between the different levels involved. Failing to provide understating (beyond mere prediction) may lead risky clinical treatment.

References

Aftab, A., Jerotic, S. (2021) Scientific Pluralism is the only way forward for Psychiatry. Acta Psychiatrica Scandinavica 2021; 143 (6): 537-538 https://doi.org/10.1111/acps.13298

Bueter, Anke (2019). A Multi-Dimensional Pluralist Response to the DSM-Controversies. Perspectives on Science 27 (2):316-343.

Cronbach, Lee J. & Meehl, P. E. (1956). Construct validity in psychological tests. In Herbert Feigl & Michael Scriven (eds.), Minnesota Studies in the Philosophy of Science. , Vol. pp. 1--174.

Cuthbert B. N. (2014). The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. World psychiatry: official journal of the World Psychiatric Association (WPA), 13(1), 28–35. https://doi.org/10.1002/wps.20087

Sullivan, J. A. 2009. "The Multiplicity of Experimental Protocols: A Challenge to Reductionist Neuroscience." Synthese 167: 511–539.

Sullivan, J. A. 2010. "Reconsidering 'Spatial Memory' and the Morris Water Maze." Synthese 177: 261–283.

E. Vintiadis (2016) 'The Importance of Pluralism in Psychiatry' (in English) in Βιοηθικοί Προβληματισμοί II Maria Kanellopoulou-Botti and Fereniki Panagopoulou (eds). Athens: Papazisis.

Mathematics as a New Way of Reasoning: The Case of Electrostatics in the 18th Century

Lucas Marcelo C. Nardi and Cibelle C. Silva

Mathematics as a New Way of Reasoning: The Case of Electrostatics in the 18th Century

Dr. Lucas Marcelo Cavalari Nardi; Prof. Dr. Cibelle Celestino Silva

Group of History, Theory and Science Teaching, University of Sao Paulo, Brazil

www.ghtc.usp.br

Long Abstract:

Mathematization has been a topic entrenched within the historiography of science (Lenhard & Carrier 2017; Ferreira & Cibelle, 2020), with the notable example of Kuhn's importance of the mathematization as a topic of research for historians of science (Kuhn, 1977). However, the process of the mathematization of the studies of natural science occurred in different fields in different historical moments and contexts, each demanding specific deeper analysis. In this paper, we analyze the process of the mathematization of electrostatics throughout the 18th century, using as a philosophical perspective the notions of styles of mathematization and epistemic projects. In the end, we show how these mathematized developments played an essential role in the old clash between ethereal theories and action-at-a-distance theories, a debate that pervaded physics up until the beginning of the 20th century.

To cement our point, we focus on the contributions of Johann Albrecht Euler, Franz Ulrich Theodosius Aepinus, and Charles-Augustin Coulomb. Johann Euler, for instance, is a prime example of a mathematized work on electrostatics based on an ethereal theory of electricity. His explanations for electrical phenomena were based on a mechanistic understanding (*i.e.*, a mechanistic picture) of the world—that is, it was based on a subtle fluid, moving, pushing, and pulling electrified bodies. Thus, Euler used notions from hydrodynamics to construct differential equations for describing the movement of the ether, understood as a fluid. The solutions for such equations added mathematical rigor for his explanations of electrical phenomena; however, the core of Johann Euler's theory was not his calculations but his mechanical reasonings with purportedly tangible entities (that is, the ether).

In contrast, Aepinus elaborated explanations of electrical phenomena based on the notion of action at a distance and mathematical calculations. Aepinus left mechanistic understandings as a backdrop to his theory. To exemplify, Aepinus' explanation for the famous Leyden jar experiment, in which an experimenter receives an intense shock after electrifying a glass jar, was summed up by a mathematical equation, something anew to electrical inquiries at the time. Similarly, Coulomb's theory of electricity also prioritized mathematical reasoning over the mechanistic one. However, unlike Aepinus, the French underplayed the roles of physical entities and mechanistic reasoning entirely by valuing only the mathematical results, which eventually led his work to physical contradictions—for instance, he deduced a theorem that was not valid for the electrical theory he was defending (the two-fluid theory). Coulomb did not address this contradiction and

implicitly claimed that his endorsed mechanistic picture was not as relevant as the mathematical result he deduced.

As argued, we advance our point using the notion of style of scientific reasoning developed by the philosopher Ian Hacking (Hacking, 2012; Bueno, 2012), a concept that aims at a philosophically understanding of how science poses the research questions that drive itself. It is a way to understand the different reasonings and techniques, with shared abstract characteristics, for proposing scientific research questions and designing research experiments and answers. We also developed the notion of epistemic projects, which refers to a combination of similar styles that constitutes a *longue durée* trajectory that can be used to analyze long historical periods of the history of science (Nardi, 2021). An epistemic project is supposed to combine similar styles (in the historical episode here discussed, styles of mathematization). From this perspective, we observe two epistemic projects coexisting along the 18th century and traversing the 19th century in studies on optics and electromagnetism—one based on the ether (where Johann Euler is situated) and another based on action at a distance (Aepinus and Coulomb). In this sense, the notion of the epistemic project is helpful since it enables the historians of science to extend their focal point of inquiry into a bigger picture in a philosophically reasonable way.

Bibliographical References:

Bueno, O. (2012). Styles of reasoning: A pluralist view. *Studies in History and Philosophy of Science , 43*(4), 657-665.

Ferreira, C. T. T.; Cibelle, C. S. (2020). The Roles of Mathematics in the History of Science: The Mathematization Thesis. *Transversal: International Journal for the Historiography of Science*, 8, 6-25.Hacking, I. (2012). 'Language, Truth and Reason' 30 years later. *Studies in History and Philosophy of Science Part A, 43*(4), 599-609.

Lenhard, J.; Carrier, M (Eds.). (2017). *Mathematics as a Tool Tracing New Roles of Mathematics in the Sciences*. Boston Studies in the Philosophy and History of Science, vol. 327. Cham: Springer.

Kuhn, T. (1977). Mathematical versus Experimental Traditions in the Development of Physical Science. *In*: Kuhn, T. *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: The University of Chicago Press, 31-65.

Nardi, L. M. C. (2021). *A matematização da eletrostática no século XVIII: de rupturas epistemológicas a estilos de matematização* [In English: *The mathematization of electrostatics in the 18th Century: from epistemological ruptures to styles of mathematization*]. Ph.D. thesis defended at the University of São Paulo.

Short Abstract:

The process of the mathematization of the studies of natural science occurred in different fields in different historical moments and contexts. In this paper, we analyze the process of mathematization of electrostatics throughout the 18th century, using as a philosophical perspective the notions of styles of mathematization and epistemic projects. In the end, we show how these mathematized developments played an important role in the old clash between ethereal theories and action-at-a-distance theories, a debate that pervaded physics up until the beginning of the 20th century.

Precisely situated individuals: Autistic ecological niche construction

Janko Nešić

Philosophy of Cognitive and Behavioral Sciences

Precisely situated individuals: Autistic ecological niche construction

Autism spectrum disorder (ASD) is a psychopathological condition characterized by persistent deficits in social interaction, social communication and restricted, repetitive patterns of behaviour and interests (APA 2013). To build an integrative, ecological-enactive account of autism, I propose we should endorse the skilled intentionality framework (SIF; Rietveld, Denys, & van Westen 2018). SIF connects a number of disciplines - ecological psychology (landscape of affordances), phenomenology (selective openness to and relevance of affordances, optimal grip), emotion psychology (states of action-readiness), and embodied neurodynamics (self-organizing affordance-related states of action-readiness). In SIF, embodied cognition is understood as skilled engagement with affordances (possibilities for action) in sociomaterial environment of the ecological niche by which an individual tends toward the optimal grip.

An important part of SIF is an ecological-enactive interpretation of the free energy principle and predictive processing framework (Bruineberg and Rietveld 2014). The predictive brain tries to minimize prediction-errors that result from (mis)matching between top-down predictions and bottom-up sensory information. Brain instantiates a hierarchical probabilistic model of the environment, the "generative model". Agent gives more or less precision to either priors beliefs or current sensory evidence (prediction errors) depending on how reliable (or "precise") they estimate each to be. In predictive processing, mental health is understood in terms of the goodness of the generative model of the agent. In SIF`s non-representational interpretation, the generative model is viewed as a multiplicity of simultaneous and coupled states of action-readiness that are sensitive to some affordances (selective openness) accessible in the landscape of affordances (Bruineberg, Kiverstein and Rietveld 2018). Predictive processing accounts point out that in ASD too much precision is assigned to prediction errors (Pellicano and Burr 2012; Van de Cruys et al. 2014; Constant et al. 2020; Miller et al. 2022). According to the HIPPEA theory ("high and inflexible estimation of precision of prediction errors", Van de Cruys et al. 2014), autistics designate atypically high precision to bottom-up prediction errors and have trouble adapting to environmental uncertainties which leads to a restricted focus in perception and demand for sameness and stereotyped behaviour, and these are strategies they resort to in order to cope with a great amount of prediction error, trying to make the sensory environment more predictable (Constant et al. 2020, 614). Autistic persons depend heavily on current sensory information and less on prior beliefs (Miller, Kiverstein, Rietveld 2022), they give too much weight to novel sensory evidence and so cannot attune to stable regularities (Kirchhoff and Kiverstein 2020; Lawson et al. 2014; Palmer et al. 2017).

Autistics suffer from suboptimal generative models that do not reach high levels of abstraction and generality (they build "overfitted" models). I will argue that in SIF`s terms, autistic patterns of action-readiness pick out very specific solicitations in the environment and achieve optimal grip only in well-known situations and specifically constructed ecological niches. They make interventions in the environment with reliable cue-effect relations. Autistics experience complex social environments as foreign and avoid natural sensory niches that cannot be reliably predicted. To reduce uncertainty, they over-rely on routinized behaviour, strict habits, sameness, and a familiar environment - a predictable ecological niche that they construct. They lack the openness (pathological embodiment) needed to be responsive to the relevant affordances, pilling up habits and skills that are rigidly applied without adjustment to the changing environment. Autistic persons favour social environments that increase predictability through ritual behaviour and routines and design monotonous landscapes of affordances.

I argue that endorsing the skilled intentionality framework, helps us understand the ecological particularities of autism spectrum disorder. Autistics suffer from suboptimal generative models that do not reach high levels of abstraction and generality (predictive processing) Autistics experience complex social environments as foreign and avoid natural sensory niches that cannot be reliably predicted. To reduce uncertainty, they over-rely on routinized behaviour, strict habits, sameness, and a predictable ecological niche that they construct. They lack the openness needed to be responsive to the relevant affordances, pilling up habits and skills that are rigidly applied without adjustment to the changing environment.

Keywords: autism, ecological niche, predictive processing, affordances, skilled intentionality

Precisely situated individuals: Autistic ecological niche construction

I will argue that endorsing the skilled intentionality framework, helps us understand the ecological particularities of autism spectrum disorder. Autistics suffer from suboptimal generative models that do not reach high levels of abstraction and generality (predictive processing) Autistics experience complex social environments as foreign and avoid natural sensory niches that cannot be reliably predicted. To reduce uncertainty, they over-rely on routinized behaviour, strict habits, sameness, and a predictable ecological niche that they construct. They lack the openness needed to be responsive to the relevant affordances, pilling up habits and skills that are rigidly applied without adjustment to the changing environment.

A dualist model about powers and laws in light of the wave function

Maria Panagiotatou

a) General Philosophy of Science

A dualist model about powers and laws in light of the wave function

Keywords: wave function, quantum mechanics, laws, powers, dualist model

Short abstract

My aim is to analyse the status of the wave function in quantum mechanics and examine the prospects of a dualist model in the metaphysics of science with laws and powers equally fundamental. I will discuss views that attribute law-like status to the wave function and views that give power-based descriptions of it. In this light, I will examine whether there is room for unifying the law-like and the power-based views about the wave function in order to suggest a better understanding of its role in the context of nonrelativistic quantum mechanics.

Extended abstract

With respect to non-Humean approaches, there is a vivid discussion in the metaphysics of science concerning the nomological structure of the world and the origin of natural necessity. There are traditional accounts of laws as the source of natural necessity (ADT-approach defended by Dretske 1977, Tooley 1977, Armstrong 1983); and, there are recent views that consider powers fixing the nomic structure of the world either without laws (Mumford & Anjum 2011) or with laws playing no fundamental role in the ontology (Bird 2007). It is common knowledge that all monistic accounts face several problems. For instance, the law-based approaches fail to give an adequate account of how laws play their governing role. As for the power-based approaches, they fail to offer an adequate account of conservation laws and symmetries in physical theories.

In a recent paper, Ioannidis, Livanios, and Psillos (2021) proposed a dualist model that treats laws and powers as equally fundamental concerning their role in the nomological structure of the world. Briefly, according to their model, the laws govern the behaviour of worldly things, and worldly things can execute the laws because of the powers they possess.

The aim of my talk is to examine the prospects of such a dualist model like the above in light of the status of the wave function in quantum mechanics. Broadly speaking, in physicists' mind, the wave function is a mathematical representation of a quantum state; in other words, the wave function of a system (individual or entangled) at some instant describes the system's state at that instant and incorporates its dynamical properties. Despite this straightforward description of the wave function, the correct way to understand it has been, since the birth of quantum mechanics, an ambiguous issue. From the statistical interpretation of the wave function in the early years of quantum mechanics till the recent, much discussed, proposal to view the wave function as a high-dimensional physical field, the efforts to comprehend the status of the wave function are many and considerably different.

I will especially look into views that attribute a law-like status to the wave function, and views that give power-based descriptions of it. Preferences of philosophers seem in some cases to depend on the version of quantum theory that a philosopher adopts. For example, Goldstein and Zanghì (2013) discuss in the context of Bohmian mechanics a nomological or a quasi-nomological role of the wave function. Allori (2017) describes how the wave function behaves more like a gauge potential than a field, and also proposes a nomological (law-like) view of the wave function instead of an ontological view. In contrast, we have Suárez (2015) who criticises the nomological interpretations of the wave function because of its dynamical nature, since laws are not supposed to change with time. Therefore, Suárez (2015) proposes a dispositionalist interpretation of the wave function in Bohmian mechanics. Dorato and Esfeld (2010) claim that the Ghirardi, Rimini, and Weber formulation of quantum mechanics (GRW theory) favours an ontology of powers or dispositions.

In light of the above, I will examine whether there is room for unifying the law-based and the power-based views about the wave function in order to give a better understanding of its role in the context of nonrelativistic quantum mechanics, and, if possible, achieve this without adopting a particular interpretation. If this is the case, I will cheque whether the dualist model of Ioannidis, Livanios, and Psillos (2021) can accommodate this possible unification.

References

Armstrong, D. M. (1983). *What is a law of nature?* Cambridge: Cambridge University Press.

Allori, V. (2017). "A New Argument for the Nomological Interpretation of the Wave function: The Galilean Group and the Classical Limit of Nonrelativistic Quantum Mechanics", *International Studies in the Philosophy of Science* 31(2), 177-188.

Bird, A. (2007). *Nature's metaphysics: Laws and Properties*. Oxford: Oxford University Press.

Dorato, M., & Esfeld, M. (2010). "GRW as an ontology of dispositions". *Studies in History and Philosophy of Modern Physics* 41(1), 41-49.

Dretske, F. (1977). "Laws of nature". *Philosophy of Science* 44, 248-268.

Goldstein, S., & Zanghì, N. (2013). "Reality and the Role of the Wave Function in Quantum

Theory". In *The Wave Function: Essays in the Metaphysics of Quantum Mechanics*, edited by David Z. Albert and Alyssa Ney, 91-109. New York: Oxford University Press.

Ioannidis, S., Livanios, V., Psillos, S. (2021). "No laws and (thin) powers in, no (governing) laws out". *European Journal for Philosophy of Science* 11(1).

Mumford, S., & Anjum, R. L. (2011). *Getting causes from powers*. Oxford: Oxford University Press.

Suárez, M., (2015). "Bohmian dispositions". *Synthese* 192 (10), 3203-3228.

Tooley, M. (1977). "The nature of law". *Canadian Journal of Philosophy* 7, 667-698.

General philosophy of science. Title: Variable relativity of causation is good

Veli-Pekka Parkkinen

a) General philosophy of science

Title: Variable relativity of causation is good

Keywords: causation; interventionism; variable relativity

Extended abstract

(675 words)

Interventionism is a theory of causation with a pragmatic goal: to define causal concepts that are useful for reasoning about how things could, in principle, be purposely manipulated. In its original presentation, Woodward's (2003) interventionist definition of causation is relativized to an analyzed variable set. Responding to criticisms that focus on this "variable relativity" (e.g. Strevens, 2007), in (Woodward, 2008) Woodward changes the definition of the concept of contributing cause, which defines the basic notion of causal relevance for interventionism, so that it is no longer relativized to a variable set. This derelativization of interventionism has not gathered much attention, presumably because it is seen as an unproblematic way to save the intuition that causal relations are objective features of the world. I argue that this move has problematic consequences. Derelativization entails two concepts of unmediated causal relation that are not coextensional, but which nonetheless do not entail different conclusions about manipulability relations within any given variable set. More specifically, in certain causal structures, a variable may be an unmediated (contributing) cause of another in the derelativized sense, and yet not be a direct cause as defined by interventionism, depending on the choice of variables included in the analyzed variable set. Nonetheless, these two concepts of unmediated causal relation never entail different manipulability claims about any set of analyzed variables. This conflicts with the core principle of interventionism, according to which every distinct (model of a) causal structure over a set of variables entails a distinct set of claims about manipulability relations among those variables, and conversely, every completely specified set of manipulability claims corresponds to a distinct causal structure: "*No causal difference without a difference in manipulability relations, and no difference in manipulability relations without a causal difference*" (Woodward, 2003, p. 61).

I then argue that there is no obvious solution to this problem that would preserve the derelativized concept of contributing causation, without creating more problems. The concept of direct causation cannot be derelativized to align with a derelativized concept of contributing causation, unless one also insists that the concept of direct causation only applies at the finest possible grain of description, which is an unrealistic and methodologically useless requirement. Excluding the problematic structures from the scope of interventionism would rule some *prima facie* causal structures as non-causal. Biting the bullet and treating the distinction as a mere technicality is not an option if one wants to preserve the pragmatic orientation of interventionism.

I conclude that interventionist causation should not be derelativized in the first place. This conclusion is rendered acceptable by pointing out that a definition of causation as manipulability is applicable only when a distinction is drawn between a system of which we ask causal questions about, and an environment from which the system can be manipulated. Causal reasoning so understood is never about the world as a whole, but about some locally defined structure of dependencies that can in principle be interfered with by exogenous interventions. The distinction between a target system and its environment is drawn by an agent that engages in causal reasoning, is influenced by the interests and background knowledge of the agent, and in the formal machinery of interventionism it amounts to a decision to focus on one variable set rather than another. How the distinction between the target system and its environment is drawn will on occasion affect what can truthfully be concluded about manipulability relations between parts of the target system. Hence, if causal concepts are to reliably track manipulability relations, causal concepts should be sensitive to how the boundaries of the target system are drawn. Given the goal that interventionism sets for itself – to describe causal concepts that are useful tools for reasoning about manipulation and control – variable relativity is a required feature of the theory, not an unwanted technical artefact. If one opposes interventionism on the grounds of variable relativity, one presumably then opposes the very idea that the primary function of causal concepts is to guide reasoning about manipulability relations. Such criticism should thus be accompanied by reflection about what alternative aims causal reasoning may serve that are more important, and how a philosophical theory of causation should address those goals.

Short abstract

(100 words)

Interventionism has a pragmatic goal: to define causal concepts that are useful for reasoning about manipulation and control. Originally, interventionism defined causation relative to a variable set (Woodward, 2003). In (Woodward, 2008), the most general interventionist notion of causation, contributing cause, is derelativized. I argue that derelativization entails two concepts of unmediated cause that are not coextensional, but do not entail different conclusions about manipulability within any variable set. This conflicts with the stated pragmatic aim of interventionism. I discuss putative solutions but find them all wanting. I conclude that interventionist causation should not be derelativized. Various considerations are offered rendering that conclusion acceptable.

References

Strevens, M. (2007). Review of Woodward,"Making Things Happen". *Philosophy and Phenomenological Research 74*(1), 233–249.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford university press.

Woodward, J. (2008). Response to Strevens. *Philosophy and Phenomenological Research 77*(1), 193–212.

Scientists as Agents of Democratization in Authoritarian Societies

Viorel Pâslaru

Scientists as Agents of Democratization in Authoritarian Societies

Short abstract
I examine an open letter that a group of scientists addressed in 1988 to the authorities of Moldavian Soviet Socialist Republic and challenged the official view of the Communist Party and the state on a politically sensitive issue regarding the name, alphabet, and status of the language spoken in that republic. I argue that the letter asserted values fundamental to democracy, offered a model of public deliberation, claimed the autonomy of science from political intrusion and its status of an autonomous branch of power, and contributed to the democratization of the Moldavian republic.

Long abstract
The interplay between science and society is usually examined in the context of Western democracies. Scholars have shed light on various aspects of democratic engagement with science, commercial and political pressures upon science, and the proper place of science in democratic societies. However, focusing on the interaction between democratic societies and science does not consider that significant amounts of research are done in anti-democratic societies and that scientists have played important roles in catalyzing democratization of some authoritarian societies. In this presentation, I examine one instance of scientists' involvement in public life that helped transition the former Moldavian Soviet Socialist Republic (MSSR) to democracy.

In September 1988, 66 intellectuals, most of whom were PhDs in Philology, Physics, Mathematics, History, Pedagogy, Psychology, addressed an open letter to the interdepartmental committee of the MSSR legislature tasked with the study of problems of history and issues of development of Moldovan language. The letter urged the committee to pursue these tasks: 1) reject the theory of two Romance languages north of Danube, and implicitly recognize that a Moldovan language does not exist; 2) re-introduce the Latin alphabet to replace the Cyrillic; 3) make the language of Moldovans an official language; 4) adopt an interdisciplinary and comprehensive set of measures to create the necessary conditions for the unrestricted functioning of the language of Moldovans; and 5) consider that those measures will diminish the tense interethnic relationships due to the subpar status of the language.

This letter catalyzed the transition to democracy in MSSR. Demands 1) and 2) challenged the official Soviet line that had asserted for decades that Moldovan is a language distinct from Romanian, although similar, and that it should use the Cyrillic alphabet. The Communist Party of MSSR held the authority to decide what was scientific and what scientists (and other citizens) could say in public, which represented a direct intrusion of the political into the scientific. By challenging the official dogma, scientists who signed the letter contested the authority of the Party to prescribe scientific truths, claimed the autonomy of science, and asserted the separation of powers, which characterizes democracies. Demands 4) through 5) raised issues of linguistic justice because the language of the indigenous majority was seen as oppressed by Russian, language of a minority in MSSR that yielded immense political, economic, and social power. That scientists challenged the Party in public was an act of asserting freedom of speech. That the editor of a newspaper agreed to publish scientists' letter was an act of defying party control and asserting a free press. Finally, by publicly challenging the official line of the state, which per Soviet Constitution was led by the Communist Party and whose leadership could not be

contested, the letter asserted equality of citizens and the state. Citizens followed suit and forced public debates on those issues, bringing about social change. The letter asserted values fundamental to democracy.

In addition to contributing to the democratization of MSSR, the letter suggests that the proper role of science in society is neither that of passive, on demand resource, nor of technocratic ruler, but an autonomous branch of power that can challenge the political when it oversteps its boundaries or does not address social issues. Additionally, the letter aligns with other efforts by Soviet and non-Soviet scientists to refuse unacceptable political intrusion into science and press politicians to uphold human rights and act on social issues.

Over the course of the following year, linguists who published the letter, as well as others who did not get a chance to sign it, published a number of articles offering linguistic, historical, and linguistic justice arguments in support of their demands stated in the letter. A public debate involving opposing views ensued. By contributing to the public debate scientific evidence and argumentation, linguists offered a model of reasonable deliberation that the nascent democracy could use.

Selected Bibliography:

Atwoli, L., Baqui, A. H., Benfield, T., Bosurgi, R., Godlee, F., Hancocks, S., ... & Vázquez, D. (2021). "Call for emergency action to limit global temperature increases, restore biodiversity, and protect health." The British Medical Journal, 374:n1734. http://dx.doi.org/10.1136/bmj.n1734

Brown, M.B., (2009). *Science in democracy: Expertise, institutions, and representation*. MIT Press: Cambridge, MA.

Collins, H., & Evans, R. (2017). *Why democracies need science*. John Wiley & Sons: Cambridge, UK.

Douglas, H., 2021. *The Rightful Place of Science: Science, Values, and Democracy: The 2016 Descartes Lectures.* Consortium for Science, Policy & Outcomes.

Gilligan, E. (2004). *Defending human rights in Russia: Sergei Kovalyov, dissident and human rights commissioner, 1969-2003*. Routledge.

Hartl, P. and Tuboly, A.T. eds., 2021. *Science, Freedom, Democracy*. Routledge: New York and London.

Jasanoff, S. (1998). *The fifth branch: science advisers as policymakers*. Harvard University Press: Cambridge, MA.

Joravsky, D. (2010). *The Lysenko affair*. University of Chicago Press: Chicago.

Istru, Bogdan, *et al*. (1988). Open Letter (Scrisoare Deschisă), Învățământul Public, September 18, p. 4.

Jimulev, I. F., & Dubinina, L. G. (2005). New Information about "The Letter of Three hundred" – a massive protest from 1955 of Soviet scientists against Lysenkoism, *Vavilov Journal of Genetics and Breeding*, *9*(1), 13-33) (Жимулев, И. Ф., & Дубинина, Л. Г. (2005). Новое о «Письме трехсот»—массовом протесте советских ученых против лысенковщины в 1955 г. *Вестник ВОГиС*, *9*(1), 13-33).

Lacey, H. (2005). *Is science value free?: Values and scientific understanding*. Routledge: London.

Pringle, P. (2008). *The murder of Nikolai Vavilov: The Story of Stalin's persecution of one of the great scientists of the twentieth century*. Simon and Schuster: New York.

Ripple, W. J., Wolf, C., Newsome, T. M., Galetti, M., Alamgir, M., Crist, E., ... & 15,364 Scientist Signatories from 184 Countries. (2017). World scientists' warning to humanity: a second notice. *BioScience*, *67*(12), 1026-1028.

Smith, H. (2020). Yuri Orlov, physicist who became a symbol of Soviet dissent, dies at 96, *Washington Post*, September 30.

Purifying applied mathematics and applying pure mathematics: How a late Wittgensteinian perspective sheds light onto the dichotomy

José Antonio Pérez Escobar and Deniz Sarikaya

Purifying applied mathematics and applying pure mathematics: How a late Wittgensteinian perspective sheds light onto the dichotomy

Track: f) Formal Philosophy of Science and Philosophy of Mathematics

Keywords: Later Wittgenstein; embodied mathematics; Philosophy of Applied Mathematics;

Short Abstract (99 Words):

We argue that there is no strong demarcation between pure and applied mathematics. We stress non-deductive components within pure mathematics based on an embodied view of mathematics and "purer" components of applied mathematics, like the theory of the models that are concerned with practical purposes. Some mathematical theories can be viewed through either a pure or applied lens. We then discuss how this view relates to different interpretations of Wittgenstein. Building on a maverick interpretation by Dawson, and endorsing an extended notion of meaning as use which includes social, mundane uses, we elaborate a fuzzy, but more realistic, demarcation.

Long abstract (670 Words)

The pure/applied distinction in mathematics is often taken for granted without much thought put into it. One of the few philosophers that have engaged with this dichotomy is Wittgenstein, especially the later Wittgenstein. The later Wittgenstein's work on philosophy of mathematics, despite its numerous and original insights, has received little attention, and hence, interpretations and applications of this work are still lacking or in process. We believe that one of the potential applications of this work is the building of a demarcation between pure and applied mathematics.

We argue that at the base of the pure/applied mathematics dichotomy, which mainly describes sociologically distinct groups, lie different community-specific priorities (regarding evaluative standards, directness of application and rhetoric) instead of a deeper metaphysical distinction between two putative realms. We show that Maddy's exegesis of the later Wittgenstein's philosophy of mathematics (Maddy, 1993) fails to capture such practice-based differences, and hence is not useful for our main aim, namely, to build a demarcation between pure mathematics and applied mathematics. However, another, maverick exegesis of the later Wittgenstein, recently proposed by Dawson (2014) and which has not received much attention yet, may be more appropriate for our aim. While this is not an exegetical work, given the principle of charity, we will endorse this maverick exegesis, further build on it, and apply it to our philosophical endeavor.

More specifically, we first draw from a classical, revisionary interpretation of Wittgenstein's late philosophy, which, we argue, cannot account for practices in pure mathematics. Instead, we propose a sociological analysis of practices based on the Wittgensteinian notion of meaning as use. We acknowledge a sociological division between pure mathematics and applied mathematics communities based on putatively extramathematical factors: there are different prizes, journals, chairs, departments, and so forth. In line with late Wittgensteinian philosophy, we believe that this sociological distinction is ultimately rooted in the different ways in which each community uses mathematics. Therefore, we analyze their uses of

mathematics to understand what is meant when one talks about applied mathematics or pure mathematics. We note that the same mathematics may be used differently in different contexts, and hence, “pureness” is not a property of the symbols or diagrams that comprise mathematics. By family resemblance, another key concept from Wittgenstein’s late philosophy some of these uses can be considered “pure” and others can be considered “applied”, but these uses, instead of being strongly demarcated as in section 2, share many traits.

Then, we discuss two aspects of pure mathematics which show that it is in fact grounded in the real world, in contrast to the tenets of Maddy’s exegesis of the late Wittgenstein. These aspects are 1) the role of abstraction or mathematization as an iterative process from the real world to pure mathematics (what we call the neo-aristotelian view) and 2) the idea that meaning from our everyday experience is transferred to the realm of pure mathematics (called embodied mathematics, cf. (Lakoff & Núñez 2000)). While it is true that pure mathematics can in principle start from arbitrary axioms, this does not reflect the actual research practice. There is a large modelling-like part when building pure theories. A mathematical theory usually starts from informal notions; it might not be entirely clear how to define the structures mathematicians talk about and there might be different competing notions or clarifications of those. Older disciplines are usually more settled and working in concrete axiomatic settings.

Finally we note how the above considerations are compatible with and able to be philosophically framed by another, maverick interpretation of Wittgenstein’s late philosophy of mathematics, one that accepts meaning in pure mathematics (Dawson, 2014) and we build on this interpretation by extending meaning as use sociologically, and further elaborate on the fuzzy demarcation of pure and applied mathematics.This new account involves a three-layered meaning as use perspective that we will unpack in our talk.

Literature:

Dawson, R. (2014). Wittgenstein on pure and applied mathematics. *Synthese*, 191(17): 131-4148.

Lakoff, G., & Núñez, R. (2000). Where mathematics comes from. New York: Basic Books.

Maddy, P. (1993). Wittgenstein’s anti-philosophy of mathematics. V. Phul (Ed): Wittgenstein’s Philosophy of Mathematics,Vienna: Verlag Hölder-Pichler-Tempsky: 52–72.

What is like to lucid dream? Lucidity as a test case for the knowledge argument

Stefan Petkov

EENPS 2022 section:

c) Philosophy of Cognitive and Behavioral Sciences

Title:

What is like to lucid dream?
Lucidity as a test case for the knowledge argument

Keywords: qualia, the knowledge argument, ability hypothesis, experimental philosophy, cognitive science and psychology of dreaming

Short Abstract

The problem of evaluating the knowledge argument and the type of qualia it entails has attracted enormous philosophical attention. However, the debates have proceeded mainly by constructing and interpreting ideal cases, deriving arguments based on them and evaluating the results using rational intuitions. Here we adopt a different strategy. We argue that the knowledge argument has at least one interpretation that renders itself easily to a empirical examination – the ability hypothesis. We identify lucid dreaming a good empirical correlate. We interpret existing studies on lucidity from cognitive science, psychology, and our own pilot study, as a corroboration the ability hypothesis.

Extended Abstract

The problem of how to define qualia and assert which mental states can possess them has received enormous attention in the literature of philosophy of mind. Presently a query on this topic in the web portal for philosophical papers (https://philpapers.org/) returns more than one thousand papers. A focal point in the qualia debates has been the knowledge argument. Introduced by Frank Jackson in 1982, it offers the by now familiar story of Mary the neuroscientist/physicist who has been confined in a black and white room. There she studies the phenomena of optics and human perception of colours. Eventually Mary learns all there is to learn about colours and colour vision, but she has never experienced seeing any non-monochrome object. Finally, she is released and sees for the first time the outside world full of colours. Jackson asks us – did she learn anything new or not?[1]

Despite the enormous interest, little consensus has been reached in evaluating the knowledge argument. The disagreement ranges from giving particular definitions of what qualia are and what

1 Jackson's original conclusion (i.e., that physicalism is false) remains highly controversial. We do not aim to settle this controversy here, instead we think that it would be more useful to present the knowledge argument in its canonical form without Jackson's conclusion. Therefore, we would refer to the knowledge argument as only the case of Mary plus the open question if she has learned something new from her new experience.

would happen to Mary, to discussions about which mental states have them, to philosophical positions that reject the existence of qualia altogether, or criticisms that dismiss the whole method of arguing using idealized cases.

Despite this heterogeneity, one unifying characteristic in the debates is their methodology. With few exceptions the discussions proceed mainly by using the analytical method of interpreting ideal cases, deriving arguments based on them and evaluating the results using rational intuitions. Whilst this has certainly helped to map the territory, and has led to rich and detailed conceptual analyses, the purely analytical approach has perhaps also contributed to the impasse in achieving any form of consensus.

Following this idea, we believe that a shift in strategy would be beneficial. Instead of using a purely speculative method, we aim to offer an empirical contribution. We take that at least one interpretation of the knowledge argument – the ability hypothesis if properly framed, renders itself to an empirical examination. In order to render the hypothesis testable, we reformulate it as the weaker non-reductive claim - obtaining a novel experience involves a gain in the ability to imagine and remember this experience and promotes a unique identification strategy. ***That identification strategy is the ability to recognize the experience by direct experiential insight and not via any of its descriptive features.***

Consequently, instead of taking the beaten path and analysing a bouquet of existing and potential criticisms, we suggest treating the ability hypothesis and Mary's case, similarly to an idealized analogy-based model. We take the idea that such models are developed, as a source of analogies that can be built between an idealized case (formulated by the model) and an empirical target.

We then suggest that at least one strong analogue to the ability hypothesis and Mary's case is readily available for both empirical and conceptual studies and this analogue is lucid dreaming – the veridical awareness of dreaming subjects that they are dreaming. Focusing on lucidity as a test case, we list the key positive, negative and neutral (yet unknown) similarities with Mary's room. We locate one key negative analogy with Mary's room that permits a promising avenue of tackling the ability hypothesis. Mary's case involves perceptional qualia (seeing colours when the experiential referent is present), whilst "the what is likeness of lucidity" is non-perceptional. Nevertheless, we argue that this could be to our advantage, because it permits us to easily discriminate between two strategies that lucid dreamers use to recognize that they are dreaming.

We review studies from psychology and neuroscience and interpret them as being consistent with the claim that dreamers sometimes trigger lucidity by recognizing a descriptive inconsistency between the content of their dreams and the ordinary waking experience, but sometimes also by a sudden direct realization of their conscious state. The later cases of dream recognition, which seem to not depend on any specific experiential content of dreams or how such content is experienced, might offer a preliminary corroboration of the idea that subjects can identify their experiential state by a direct insight. As such these empirical evidences can serve also as a preliminary corroboration of the ability hypothesis (as we have formulated it).

However, since the studies we discuss were not conducted with a focus on providing empirical data for the qualia debates, we find them insufficient to clearly support our claims. Therefore, we further extend and clarify on them by our own small scale pilot study. In our study, we interviewed eleven experienced lucid dreamers and used thematic analysis to analyse the interview data. Consequently, we believe that our results in combination with the existing studies provide a stronger empirical support to the ability hypothesis.

Incorporating (variational) free energy models into mechanisms: the case of Bayesian predictive processing

Michał Piekarski

Incorporating (variational) free energy models into mechanisms: the case of Bayesian predictive processing

Keyword: Predictive processing; Free energy principle; Constraints; Mechanisms; Explanation.

Long abstract:

There is a view emerging in the philosophy of science that research practices in science can be characterized in terms of discovering and describing mechanisms. Mechanistic explanations are based on the identification of mechanisms and strategies understood as their decomposition. Recently, there has been a discussion among mechanists about the necessity to include constraints and free energy flows into the explanations, as constitutive components of mechanistic explanations. This is directly related to the existence of control mechanisms that are non-autonomous and entail the existence of heterarchical networks. I refer to this as the 'constrained mechanisms approach'. My presentation examines the extent to which this approach can be applied to the predictive processing framework, which is now an influential process theory, offering a computational description of perceptual and cognitive mechanisms in terms of hierarchical generative models approximating Bayesian inference. I will argue that predictive processing models based on the free energy principle are amenable to this approach.

The presented approach is a rough framework of how to integrate PP with FEP using constraints and free energy flows. If this approach is valid, then it has certain consequences for a number of discussions among PP and FEP researchers. First of all, it allows for a new way to approaching the PP-FEP relationship. If FEP refers to self-organizing adaptive systems, as described in the dynamical system theory, that are at nonequilibrium steady-state (NESS) with their environment, then with the appropriate interpretation of the notion of mechanism, dynamical FEP models may in fact turn out to be descriptions of mechanisms: "dynamical models and dynamical analyses may be involved in both covering law and mechanistic explanations—what matters is not that dynamical models are used, but how they are used" (Zednik, 2008, 1459).

The proposed mechanistic integration of PP with FEP reveals that FEP is a normative theory for PP in the sense that it sets a norm that should be met by mechanistically non-trivial PP models, assuming the implementation of the constrained mechanisms approach and its heuristics. According to this norm, PP models should have an energetic component if they are

to be mechanistic. This approach can be treated as a voice in the discussion on the status of PP and its relation to the FEP. In this approach, FEP not only constraints the space of possible algorithms for PP (Spratling, 2017), but also indicates energetic constraints for the causal organization of all autonomous systems, including those that are armed with generative models and are or should be the subject of (mechanistic) explanations formulated on the basis of PP.

In practice, this means that all autonomous systems that can be described in terms of (Bayesian) generative models realizing updating priors and likelihood based on (average) prediction error should be treated *as if* they approximate Bayesian inference constrained by VFE. In other words: FEP offers a normative framework for the PP process theory, and that the PP explains the (biologically reliable) implementation of the FEP in terms of hierarchical and heterarchical active mechanisms that implement the (Bayesian) generative model.

Short abstract:

Recently, there has been a discussion among mechanists about the necessity to include constraints and free energy flows into the explanations. According to this approach, there are a number of cognitive mechanisms that cannot be satisfactorily explained by using decomposition as understood in traditional way.

In my presentation, I will consider how this approach can be applied to predictive processing, where control mechanisms play an extremely important role. Predictive processing is a new process theory of the brain that provides a computational model of cognitive mechanisms. I will argue that predictive processing models based on free energy principle are amenable to “constrained mechanisms” approach.

References:

Bechtel, W. (2019). Resituating cognitive mechanisms within heterarchical networks controlling physiology and behavior. Theory & Psychology, 29(5), 620–639. https://doi.org/10.1177/0959354319873725.2020.

Bechtel, W. & Bollhagen., A. (2021). Active biological mechanisms: transforming energy into motion in molecular motors. Synthese, 1-25. https://doi.org/10.1007/s11229-021-03350-x.

Cumming, G. S. (2016). Heterarchies: Reconciling Networks and Hierarchies. *Trends Ecol Evol*, 31(8), 622-632. https://doi.org/10.1016/j.tree.2016.04.009.

Friston, K. J. (2019). A free energy principle for a particular physics. arXiv 2019, arXiv:1906.10184.

Friston, K. J., FitzGerald, T., Rigoli, F., Schwartenbeck P. & Pezzulo, G. (2017). Active inference: A process theory. *Neural Computation*, 29(1), 1–49.

Hohwy, J. (2015). The neural organ explains the mind. In T. Metzinger & J. M. Windt (Eds.), *Open MIND*, 19(T), (pp. 1–22). Frankfurt am Main: MIND Group. https://doi.org/10.15502/9783958570016.

Hooker, C. A. (2013). On the import of constraints in complex dynamical systems. *Foundations of Science*, *18*(4), 757–780. https://doi.org/10.1007/s10699-012-9304-9.

Parr, T., Da Costa, L., & Friston, K. J. (2020). Markov blankets, information geometry and stochastic thermodynamics. Philosophical Transactions of the Royal Society A, 378(2164), 20190159.

Pattee, H. H. (1972). Laws and constraints, symbols and languages. In C. H. Waddington (Ed.), Towards a theoretical biology, Vol. 4 (248–258). Edinburgh: Edinburgh University Press.

Zednik, C. (2008). Dynamical models and mechanistic explanations. In B. C. Love, K. McRae, V. M. Sloutsky (Eds.), Proceedings of the 30th Annual Conference of the Cognitive Science Society, (1454–1459). Austin: Cognitive Science Society.

Bridging the Gap between Epistemology and Ethics through Local Knowledge: The Case of Ethnopsychiatry

Elena Popa

Section b) Philosophy of Natural Science

Bridging the Gap between Epistemology and Ethics through Local Knowledge: The Case of Ethnopsychiatry

Keywords: causality; ethnopsychiatry; interventionism; mechanisms; evidence-based medicine

While causality has been investigated in relation to evidence-based medicine, thus far has little work on causation in connection to the plurality of ontologies employed in local approaches. This paper draws on research on causality in psychiatry to provide a conceptual toolkit for expanding the discussion to the context of ethnopsychiatry. Particularly, I will look at how the interventionist and the mechanistic models work in this case, also employing recent research on mechanisms in the social sciences within a biopsychosocial setting. This contribution can help address worries about the ethical effects of the neglect of local knowledge.

Extended abstract

The relation between epistemology and ethics is apparent in research meant to shape global policies. Particularly, insufficient attention to the local context can lead to ineffective policies (Cartwright 2010). One solution to this is to revise scientific epistemology such as to pay closer attention to local conditions and include local perspectives, integrating local knowledge and scientific knowledge. Doing so, however, raises several problems, and my focus will be on epistemic ones, regarding whether current scientific knowledge and methods are at all compatible with local approaches. By local knowledge, I mean knowledge acquired by local people, in forms that may differ from how scientific knowledge is presented. This paper will look at ways of incorporating local knowledge within broader medical knowledge, particularly causal knowledge. This is important because causal connections are essential for determining which interventions work.

Current work on ethnobiology and its place within scientific knowledge has used a 'partially overlapping ontologies' (Harding 1998) framework to explain local knowledge systems (Ludwig 2018; Ludwig & Weiskopf 2019). The focus of this has been mainly ecological knowledge, with other areas emerging, such as psychiatric classification (Popa 2020). Studies of ethnobiological knowledge systems reveal that causal connections and mechanisms are part of such knowledge systems (Ludwig & El-Hani 2020), so causal claims can be investigated in relation to local knowledge. This paper will look at the structure of causal knowledge in ethnopsychiatry and its potential connection to causal explanation in psychiatry. This can help build common ground between ethnopsychiatric explanations and other types of explanations in psychiatry within a pluralistic setting. This is needed because global initiatives have been criticized for marginalizing local approaches (Mills & Fernando 2014), and pluralism has been shown to yield better results that exclusive focus on biomedical approaches (Halliburton 2020). This research will contribute to developing the framework for such pluralistic view.

This paper will use two important approaches to causality: the interventionist (Woodward 2005) and the mechanistic one (Machamer, Darden & Craver 2000). Both have been used in relation to psychiatry (Kendler & Campbell 2009; Kendler et al. 2011), but not in relation to local psychiatric knowledge. This analysis will show that some of the issues regarding causality and evidence-based medicine raised previously by Russo & Williamson (2007) among others also apply when investigating causal claims in cross-cultural contexts. I will also draw on relevant work on mechanisms and explanation in the social sciences (Shan & Williamson 2021). This has the advantage of accounting for the social aspects that local interventions typically use. Furthermore, given the importance of social determinants of mental illness, exploration in this context may yield conclusions that may be further used in biopsychosocial approaches. Overall, sketching out how causal claims featuring local approaches are possible can help single out the kind of evidence needed to investigate these claims, and help incorporate approaches that work into medicine for more effective global interventions.

References

Cartwright, N. (2010). Will this policy work for you? Predicting effectiveness better: How philosophy helps (Presidential Address). In *Philosophy of Science Association*.

Halliburton, M. (2020). Hegemony versus pluralism: Ayurveda and the Movement for Global Mental Health. *Anthropology & Medicine*, 1-18.

Harding, S. G. (1998). *Is science multicultural?: Postcolonialisms, feminisms, and epistemologies*. Indiana University Press.

Kendler, K. S., & Campbell, J. (2009). Interventionist causal models in psychiatry: repositioning the mind–body problem. *Psychological medicine*, *39*(6), 881-887.

Kendler, K. S., Zachar, P., & Craver, C. (2011). What kinds of things are psychiatric disorders?. *Psychological medicine*, *41*(6), 1143-1150.

Ludwig, D. (2018). Revamping the Metaphysics of Ethnobiological Classification. *Current anthropology: A world journal of the sciences of man*, (4), 415-438.

Ludwig, D., & El-Hani, C. N. (2020). Philosophy of ethnobiology: understanding knowledge integration and its limitations. *Journal of Ethnobiology*, *40*(1), 3-20.

Ludwig, D., & Weiskopf, D. A. (2019). Ethnoontology: Ways of world building across cultures. *Philosophy Compass*, e12621.

Machamer, P., Darden L., & Craver, C. (2000) *Thinking about Mechanisms*, Phil Sci 67: 1-25.

Popa, E. (2020). Mental health, normativity, and local knowledge in global perspective. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *84*, 101334.

Mills, C., & Fernando, S. (2014). Globalising Mental Health or Pathologising the Global South? Mapping the Ethics, Theory and Practice of Global Mental Health. *Disability and the Global South*, *1*(2), 188-202.

Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. *International studies in the philosophy of science*, *21*(2), 157-170.

Shan, Y., & Williamson, J. (2021). Applying Evidential Pluralism to the social sciences. *European Journal for Philosophy of Science*, *11*(4), 1-27.

Woodward, J. (2005). *Making things happen: A theory of causal explanation*. Oxford university press.

The role of agriculture in the rise and development of classical genetics

Marcos Rodrigues da Silva

The role of agriculture in the rise and development of classical genetics

Section (a)

Keywords: history of biology; history of genetics; constructivism

Short Abstract: According to the traditional philosophy of science, successful scientific achievements are accepted because they fit some epistemological criteria; in a constructivist approach, epistemological criteria are also fundamental for accounting acceptance, however, they are joined with social factors. The inclusion of social factors has been criticized by the traditional philosophy of science, since such inclusion would disqualify science. The constructivist reply to traditional philosophy of science is that this inclusion of social factors makes science more reliable. An example would be the rise of classical genetics in America.

Extend Abstract: Why, when scientist accept scientific achievements (theories, hypotheses, experiments, concepts, entities and so on), do they consider them as successful ones? According to the traditional philosophy of science, because these achievements have empirical-theoretical strength, consistency, coherence, explanatory power, relationship with consolidated research programs and paradigms, problem solving capacity, relationship with background knowledge, mathematical elegance and so on.

The emergence, in the 1970s, of constructivist approaches, provided a further working mechanism beyond the epistemological virtues: theories are successful also because they are successful for community: they serve the interests of members of the scientific community and are socially legitimized. It is interesting to note that, from a constructivist point of view, the philosophical explanations of the traditional philosophy of science and the constructivist conception are not exclusive, but complementary each other. Such complementarity is easy to be found in the constructivist literature, in which one can notice that both epistemological criteria and social factors must be taken into account in order to explain the acceptance of a successful scientific achievement.

However, this is not the case for traditional philosophy of science; for it, constructivism is unwelcoming not because it inserts social factors, but because it deals with social factors as cognitively significant. Traditional approaches in philosophy of science don't deny the existence of a social aspect on science; but this acknowledgment doesn't imply to recognize a worthiness concerning social factors as explanatory for understanding the acceptance of a scientific production. Furthermore, traditional approaches keep going: constructivism not only allows social factors – it would evaluate that such factors would be even more meaningful than epistemological criteria; in other words, constructivism ranks social factors ahead epistemological criteria.

For traditional approaches in philosophy of science, however, the worst is still coming: allowing social factors means to disqualify science at all. As long as science is explained either only by social factors or by social factors which are more meaningful than epistemological criteria, therefore our scientific achievements cannot be seeing as approximations to reality, but merely constructions by groups of people who politically lead the scientific community; so, science must be explained only by epistemological criteria.

This way of analyzing scientific achievements – employing traditional approaches – has two methodological guidelines. First, the focus of analysis restricts itself to scientific theories as something *already consolidated*. However, it is possible to use another methodological guideline: analyzing how the scientific production took place (its development, so to speak) – and thereby to understand, for instance, socio-community interactions that drive scientific research as much as traditional epistemological categories do. Second, traditional approaches regard the statements of scientific theories as significant units of philosophical analysis; however, there are other relevant units, and one of them is exactly what is called "scientific practice": the gathering of papers, experimentation, scientific instruments, theorizing and institutionalizing of a scientific theme – institutionalization that also arises through the gathering of scientific knowledge with social institutions interested in the development of a particular science.

I'm going to apply the constructivist model to understand the rise and development of classical genetics in America, which was underlain both on the theoretical and experimental achievements of Mendelism and on some interplays of scientists among themselves and between the community understood in a broad sense (institutional). On the one hand, classical genetics has emerged and has developed by overcoming (in some extent) the explanatory models of heredity of the 19th century, carrying out controlled experimentation, searching for information about biological processes, attempting to establish a gene theory and so on; on the other hand, it had been arisen and had been developed due to the practical needs of agriculture, financial support from institutions and governments and even promises about the usefulness of genetics itself and so on. It is useful to notice, however – and this presentation will seek to develop this approach of philosophical research –, that the acknowledgment of social factors (such as those indicated above) does not take place by means a downgrading of genetics; conversely, the acknowledgment of such factors was indeed a key point for the development of genetics. Thus, an analysis of the consolidation of genetics can take into account both epistemological criteria and social factors.

In this presentation, the social factor that will be taken into account is the role of agriculture for the rise and development of classical genetics. Several historians point out that there was a demand from breeders' associations for the development of reliable techniques for the emergence of more profitable commodities from a financial point of view. This demand, in turn, was undertaken by the United States Department of Agriculture and by research funding agencies, such as the Carnegie Institution of Washington.

The first part establishes the conceptual meaning of the constructivist program that I'm going to employ in this presentation. The second one deals with the changes of the theoretical model of heredity studies from 1900 onwards. The third part introduces, from the historiography, the significance of agriculture for the emergence and development of classical genetics in America. Finally, I put forward a philosophical explanation – from constructivist references – about the consolidation of classical genetics in America.

Understanding the Internet

László Ropolyi

Contributed paper to the section "General Philosophy of Science"

Understanding the Internet

Keywords: internet research, internet science, internet philosophy, methodological considerations

The appearance and the extended use of the internet can probably be considered as the most significant development of the 20th century. However, this becomes evident if and only if the internet is not simply conceived as a network of interconnected computers or a new communication tool, but as a new, highly complex artificial being with a mostly unknown nature. An unavoidable task of our age is to study and understand the internet, including all the things, relations, and processes contributing to its nature and use.

Studying the question what the internet is, its history provides a praxis-oriented answer. Based on the demands of the 1960s, networks of interconnected computers were built up, and for the 1980s a worldwide network of computers, the *net*, emerged and became widely used. From the 1990s the network of web pages, the world wide *web*, has been built on the net. Using the possibilities provided by the coexisting net and web, *social networks* (such as Facebook) have been created since the 2000s. Nowadays, networking of connected physical vehicles, the emergence of the internet of things, the *IoT*, seems to be an essential development. Besides these networks there is an activity to form *sharing networks* to share 'contents' (material and intellectual property, products, knowledge, services, human abilities). Currently, from a practical point of view, the internet can essentially be identified as a complex being formed from five kinds of intertwined coexisting networks: the net, the web, the social networks, the IoT, and the sharing networks. It is reasonable to seek for a theoretical *description of this complex being* instead of one or other parts of it.

This complex of intertwined coexisting and interacting networks shaped by experts and active users in the changing social and cultural environments of the late Modern Age. Because the ubiquity of the internet it is necessary to deploy the full methodological arsenal we have in favor of the understanding of the internet.

Methodological considerations – trends in internet research. Research on the internet had already been initiated at the time of the emergence of the internet. In the beginning, most research was performed in the context of informatics, computer sciences, (social) cybernetics, information sciences and information society, but from the 1990s a more specific research field, "internet research," started to form, incorporating additional ideas and methodologies from communication-, media-, social-, and human sciences. From the 2000s, internet research can be considered as an almost established new (trans-, inter-, or multidisciplinary) research field. The new discipline faced serious methodological difficulties.

Four different approaches to internet research have emerged in the last three decades:

a*) Modern scientific approach*. In this kind of research, the main deal is accepting the validity of an established (modern) scientific discipline to apply its methodology on the internet and internet use. In this way the internet or internet use can be described from computational, information technological, sociological, psychological, historical, anthropological, cognitive, etc., points of view. Such research is necessarily insensitive to the characteristics of the subject matter outside of their disciplinary fields due to the conceptual apparatus and the methodology of the selected scientific discipline, in this case to the specificity of the internet and internet use. Outcomes of these studies can be considered as specific (internet related) disciplinary statements of which the significance on the specificity of the internet is not obvious at all.

b*) Postmodern studies approach*: elaborating and applying a pluralist postmodern methodology of the so-called studies. Studies include concrete, but case by case potentially different mixtures of disciplinary concepts and methodologies that are being applied to describe the selected topic. Application of studies (e.g., internet studies, cultural studies, STS, etc.) methodology results in the creation of a huge number of relevant but separated and necessarily unrelated facts. Most research published in studies are well informed on the specificities of the internet, so the selected methodological versions in the different studies can fit well to a specific characteristic of the internet or internet use, but the methodological plurality of the studies prevents reaching any generalized, universally valid knowledge of the internet.

c) *Internet science approach* to the internet and/or internet use. here is a lack of consensus regarding how to best describe the internet theoretically, i.e., whether it is a (scientific) theory or rather a philosophy of the internet that is needed. Scientific theories on the internet presuppose that the internet is an independent entity of our world and seek for its specific theoretical understanding and description. They usually combine the methodological and conceptual apparatus of social-scientific (sociology, psychology, political theory, law, political economy, anthropology, etc.), scientific, mathematical, and engineering (theory of networks, theory of information, computing, etc.) disciplines to create a proper “internet scientific” conceptual framework and methodology.

However, there is no consensus about the fundamental specificities of the internet: the foundational principles on the nature of the internet are essentially diverse ones – and in many cases they are naïve, unconsciously accepted, non-reflective, uncertain, or vague presuppositions. Philosophical considerations can usefully contribute to overcoming these difficulties. This situation is practically the same as we have (or had) in cases of emergence of any kind of sciences: the subject matter and the foundational principles of a scientific discipline are coming from philosophical considerations.

d) *Philosophy of the internet approach*. Like the internet science, philosophy of the internet also provides a theoretical description of the internet. However, if we want to construct an internet science, we need a philosophical understanding of the internet prior to the scientific one. What is the internet? What are its most fundamental specificities and characteristics? What are the interrelationships between the internet and all the other beings of our world? Only the philosophical analyses can provide authentic answers to these questions: suggesting an understanding of “the internet as the internet”, a theoretical description of its very nature, as a totality of its all aspects, as a whole entity.

The Trace of Non-Mathematical Ancient Greek Thought in the Islamic Arithmetic Works

Fatima Saadatmand

History and Philosophy of Mathematics
e) History, Philosophy and Social Studies of Science
f) Formal Philosophy of Science and Philosophy of Mathematics

The Trace of Non-Mathematical Ancient Greek Thought in the Islamic Arithmetic Works

Fatima Saadatmand[1]

The impact of Greek thought on Islamic philosophy, astronomy, geometry, medicine, and suchlike is undeniable, but talking about a hidden and really interesting influence of Greek thought and philosophy on spiritual aspects of arithmetic is something new or at least something that hasn't been discussed so much yet. The present study deals with impressions of mystical attitude from the Greeks, which lead us to understand mathematical texts better. Arithmetic as it word presents, drives from αριθμός (means number in Greek) deals with numbers and their features. The arithmetic theorem is very closely connected with the musical theorem of ancient people, especially those who desired to understand harmony and celestial music. When Plato says the music occupies the fifth rung (in the study of mathematics and after having assigned the fourth rung to astronomy), he speaks of celestial music which results from the movement, the order, and the concert of stars that travel in space. Thus it's absolutely normal if the most of effects of Greek thought would be based on numbers and then we can claim the origin of such thought in Islamic traditions also lies in the concepts of numbers. To know the aspect of the philosophical part of this discipline, this paper focuses on numbers and their roles. In another word, the definition of numbers in the Islamic sources was greatly under the impression of Pythagoreans' faith, however, it's not of a mathematical basis, it created a new perspective on number theory as the arithmetical basis. To accomplish this aim, we seek the roots of the thoughts of the Pythagoreans and Neo-Pythagoreans and their effects on one of the most important arithmetic works of the Islamic period, namely Kitāb al-tafhīm li-awā'il sīnā'at al-tanjīm or the Book of Instruction in the Elements of the Art of Astrology in both Arabic and Persian version is one of the most important early works of arithmetic science written by Bīrūnī.
All texts are extracted from Greek sources and examine the relationship between religious themes and mathematics. "One" as the origin for producing numbers plays an important role in arithmetic texts but what makes it so special is "one is not a number" based on mathematical scholars and in most arithmetic sources. At first glance, it seems strange to regard contemporary number theory, since one is a number and also is a member of mathematical sets and series such

[1] MA Researcher, Institute for the History of Science, University of Tehran, Iran; saadatmand88@ut.ac.ir

as integers, natural numbers, odd numbers, and suchlike. On the other hand, it should deal with philosophical and mystical beliefs, one as a unit or unity which means God and good in Pythagoreans' belief, impressed the definition of number, which lies in reconstructing the Pythagorean numbers. Moreover, the view about complete numbers as well as the topics of decimal place value forms different traditions in arithmetic. So the following questions may arise: Can philosophical thought or notion in connection with mysticism be the source of inspiration for mathematical definition?! And what has been the impact of the monotheism of Muslim scholars on the way of describing the number "One"? for example in the view of "One and the unity of existence and production of other numbers from one and their return to one" which is more linked to Qur'anic verses.

KEYWORDS

Arithmetic, Philosophical Thoughts, Platonic, Pythagorean, Euclidean, Number Theory, Islamic period, Bīrūnī.

Mechanisms are insufficient for explanation

Abel Sagodi and Léon de Bruin

Philosophy of Natural Science

Mechanisms are insufficient for explanation

Keywords: mechanistic explanations, dynamical systems, neuroscience

April 7, 2022

Extended Abstract

Classifying or characterizing explanations in the special sciences has received a lot of attention recently. Kaplan and Craver claim that all explanations in neuroscience appeal to mechanisms (Craver, 2007; Kaplan, 2011). In Craver's account, in a mechanistic explanation some phenomenon is explains by revealing its underlying mechanism, which consists in "a set of entities and activities organized such that they exhibit the phenomenon to be explained" (Craver, 2007, p. 5). The entities are the components of a mechanism we consider, such as, axon terminals, vesicles, calcium ion channels, membranes and so on. However, in the context of (objective) causal relationships, they are considered to be (abstract) variables. Activities are the causal components in a mechanism. They are considered both as intrinsic dynamics of a certain entity (Kaplan, 2015) and as causal relationships between pairs of entities (Craver, 2007). Craver motivates the causal-mechanistic approach by referring to the ontic approach (Craver, 2007, p.27).

The critique of Craver of the dependencies between variables can be phrased as that the HH model (derived from other principles like the Nernst equation and some ad hoc assumptions) is not accurately describing the true (causal) relationship between the variables involved. In this reading, the critique of Craver of the HH model for the action potential boils down to empirical adequacy shortcomings. In that case, we show that we cannot rely only on interventions (if we also do not limit our scientific interests, e.g. limiting ourselves to discrete categories of features) without ending up with multiple models that are equally empirically adequate.

The examples of features of the action potential discussed (in depth) in the works of Craver and Kaplan all contain binary, discrete variables. The same can be said for the other relevant capacities of mechanisms (relevant conditions and its byproducts): they can all be considered as having two possibilities, being there or not. This restriction is necessary for the contrastive framework that Craver develops. However, when variables that can take multiple or a continuum of values are considered, contrasting between different values becomes impossible and "objective" assessments of explanations becomes impossible.

In this characterization, how Craver treats the concept of the "form" for the action potential[1] The correct way to classify the different forms of the action potential (be it visually or otherwise) is however never specified. A mechanistic-causal account of explanations which includes continuous variables is as yet to be constructed. We will discuss some possible solutions to these problems and the consequences of such solutions to the scope of mechanistic explanations. The mechanistic (contrastive) view considers a monolithic account of what can count as an aspect and relevance: only difference-making changes, all determined through (binary) contrasts.

Proponents of the mechanistic framework of explanation talk about a phenomenon (the *explanandum*) that is explained by a mechanism (the *explanans*). First, we discuss issues with identifying the causal relationship for arbitrary pairs of variables (not just binary), second, extending the condition in which the relation is identified to the background conditions relevant to under which the mechanism operates, third, the additional problems with situations where more than two and finally, freely considering temporal structure as explanans.

This narrow scope of difference-making for binary variables is where mechanistic explanations can work. Difference-making for continuous variables becomes much more troublesome and becomes near impossible for infinite dimensional phenomena without shifting attention from some unjustified assumption for what counts as a variable towards a (data/surface) model based exploration of what variables are (and could be).

We consider the case where causal relationships between two variables in a system of more than two variables have to be determined by ideal interventions, which can be read as a further elaboration on the a critique of the modularity requirement for mechanisms (Cartwright, 2004; Menzies, 2012). Craver (2007) claims on page 95 that "[i]n the context of a given request for explanation, the relationship between X and Y is explanatory if it is invariant under the conditions (W) that are relevant in that explanatory context." However, the conditions that are needed for ideal interventions (the causal graphs needs to be acyclic) are not always the same as conditions that are relevant in the explanatory context.

The only role given to dynamical system theory by Kaplan (2015) in the HH model, is that it describes the temporal organization of activity in neural systems (and hence should be considered as part of the explanans). As Kaplan (2015) correctly observes, the HH model contributes to the explanation of the action potential by describing the temporal structure. But by considering this structure as explanans leads to a situation where the explanans overlaps with the explanandum phenomenon. In the case of the action potential, the different pairwise causal relationships have to be synthesized to determine the temporal values of the various entities. Even though a part of the explanans is causal-mechanistic, the explanandum is only reached through results from dynamical systems theory, which then functions as another part of the explanans.

The mechanical analysis of Craver and Kaplan provides a useful descriptive

[1]Craver does not criticize this term, which we interpret that he, together with Kaplan, accepts it.

framework for characterizing the model variables and (some of) their interactions. Yet, without the dynamical analysis, an account of the component parts that implement the dynamics describes only part of the explanation. This all makes the causal-mechanistic framework a insufficient for a full picture of explanation in neuroscience.

To the above issues, one solution could be to relax the requirements to expect causal relationships as determined by ideal interventions for every pair of variables. Instead, surface models can be introduced either from the generalization of an ideal intervention in a certain condition or from other means through the introduction of non-mechanistic explananda. Such generalizations will need to rely on *ceteris paribus* laws to be able to cover the background conditions in which the system operates that are not covered by the ideal interventions.

Short Abstract

Classifying or characterizing explanations in the special sciences has received a lot of attention recently. Kaplan and Craver claim that all explanations in neuroscience appeal to mechanisms (Craver, 2007; Kaplan, 2011). In this article, after a short description of the mechanistic account of Craver and Kaplan and the Hodgkin-Huxley model, we will first discuss the incomplete mechanistic view of what should be considered a full characterization of the explanandum phenomenon. Furthermore, we point out some of the shortcomings of the mechanistic account of explanation in relation to only considering contrasting and difference making as necessary to determine causal relationships between variables.

References

Cartwright, N. (2004). Causation: One word, many things. *Philosophy of Science*, 71(5):805–819.

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience.* Oxford University Press.

Kaplan, D. M. (2011). Explanation and description in computational neuroscience. *Synthese*, 183(3):339.

Kaplan, D. M. (2015). Moving parts: the natural alliance between dynamical and mechanistic modeling approaches. *Biology and Philosophy*, 30(6):757–786.

Menzies, P. (2012). The causal structure of mechanisms. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(4):796–805.

The two-stage view of theory assessment, re-assessed

Samuel Schindler

Section a)

The two-stage view of theory assessment, re-assessed

Keywords: theory confirmation, theoretical virtues, historical cases, evidential uncertainty, theory-choice

Short abstract

Theory confirmation is thought to proceed in two stages: first, theories are probed for their empirical accuracy, and second, theories are then assessed for their explanatory virtues. These stages are neatly distinguished and presumed not to interact in many contexts in the philosophy of science. In this paper I will challenge this assumption.

Long abstract

It is a common view in the philosophy of science that theories are assessed in two stages: first, scientists check whether a theory is empirically adequate, i.e., if what it says about the (observable) phenomena is correct. Second, once scientists are done with the first stage, and if there is more than one theory consistent with the phenomena, scientists then compare those competitor theories on the basis of those theories' other virtues, such as simplicity, unifying power, explanatory power, external consistency, etc. (also known as "theoretical virtues" (Kuhn 1977, Schindler 2018)). I call this view the *two-stage view of theory assessment*, or simply ToSTA. I believe that ToSTA does not do justice to the complexity of scientific practice: it too neatly distinguishes two stages that are often run together in actual theory assessment – for better or worse.

ToSTA is a view that can be found in many corners of the philosophy of science. For example, ToSTA is assumed in the "inference to the best explanation": one compares the hypotheses that explain the phenomena in question and then infers the likely truth of the hypothesis that explains the phenomena "best" (Harman 1965, Lipton 2004). ToSTA is also assumed in debates about the underdetermination of theories by evidence (UTE). Realists try to break UTE by assigning epistemic value to a theory's other virtues (such as simplicity); antirealists resist such maneuvers by arguing that other virtues than empirical adequacy are merely pragmatic (van Fraassen 1980). Yet both realists and antirealists agree that any set of theories must first be assessed on the basis of the relevant evidence *before* other theoretical concerns come into play (Psillos 1999, Tulodziecki 2012).

The most popular theory of confirmation, Bayesianism, has a strong focus on the first step of ToSTA: Bayes' theorem – the holy grail of Bayesians – relates prior and posterior probabilities of hypotheses in the light of some evidence, the probability of the evidence, and the likelihood of evidence conditional on the hypothesis in question. There is little room in the formalism dedicated to the weighting of theoretical virtues, such as simplicity, in the assessment of a theory. Some Bayesians have suggested that a simple theory might be thought to have a higher *prior* probability of being true than a theory that is more complex (Salmon 1990, Earman 1992). But since many Bayesians believe that the priors "wash out" with just enough revisits of the evidence, the role Bayesians assign to theoretical virtues in the assessment of theories couldn't be more minimal (for further criticism see also Forster (1995)). Although there are attempts to formalize the unifying power of a theory in terms of likelihoods (McGrew 2003, Myrvold 2003), these proposals have been criticized on several fronts (Lange 2004, Schupbach 2011, Glymour 2014).

ToSTA is not unopposed among philosophers of science. For example, Forster and Sober (1994), in a widely-cited paper, have argued that simpler curve-fitting models tend to be preferable over more complex ones (whereby relative simplicity is measured in terms of the number of free adjustable parameters). That is so, because models that fully fit the data risk overfitting the data (i.e., they accommodate noise in the data). Simpler models therefore tend to fit future data better than more complex ones (see also Hitchcock and Sober 2004). This proposal implies that the two stages of ToSTA are perhaps not as neatly distinct as standard accounts of confirmation would have it. At least when it comes to cure-fitting. It is less clear if or to what extent Forster, Sober, and Hitchcock's conclusions carry over to higher level theories, i.e., theories that form the centre of interest in the realism debate, in which theoretical virtues have figured perhaps most prominently (Schindler 2018).

In this paper, I too want to challenge the strict separation ToSTA envisages between the two stages of theory assessment. I will do so by considering a number of historical case studies of important scientific discovery (from different sciences) in which the assessment of the theory's fit with the evidence went hand-in-hand with the assessment of a theory's virtues. In all of these cases, there was epistemic uncertainty about the reliability of the relevant evidence, but the theoretical virtues of the assessed theories boosted confidence in those theories and also helped scientists to disambiguate a confusing evidential situation.

References

Earman, John. 1992. *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*: Cambridge, ma: MIT Press.

Forster, Malcolm R. 1995. Bayes and Bust: Simplicity as a Problem for a Probabilist's Approach to Confirmation. *The British Journal for the Philosophy of Science*, **46** (3): 399-424.

Forster, Malcolm and Elliott Sober. 1994. How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *The British Journal for the Philosophy of Science*, **45** (1): 1-35.

Glymour, Clark. 2014. Probability and the Explanatory Virtues. *The British Journal for the Philosophy of Science*, **66** (3): 591-604.

Harman, Gilbert H. 1965. The inference to the best explanation. *The Philosophical Review*, **74** (1): 88-95.

Hitchcock, Christopher and Elliott Sober. 2004. Prediction versus accommodation and the risk of overfitting. *The British journal for the philosophy of science*, **55** (1): 1-34.

Kuhn, Thomas S. 1977. Objectivity, Value Judgment, and Theory Choice. In *The Essential Tension*, Chicago: University of Chicago Press, 320-333.

Lange, Marc. 2004. Bayesianism and unification: A reply to Wayne Myrvold. *Philosophy of Science*, **71** (2): 205-215.

Lipton, Peter. 2004. *Inference to the Best Explanation*. 2 ed. London: Routledge.

McGrew, Timothy. 2003. Confirmation, Heuristics, and Explanatory Reasoning. *British Journal for the philosophy of science*, **54** (4):

Myrvold, Wayne. 2003. A Bayesian Account of the Virtue of Unification. *Philosophy of Science*, **70**: 399 - 423.

Psillos, Stathis. 1999. *Scientific Realism: How Science Tracks Truth*. London: Routledge.

Salmon, Wesley. 1990. Rationality and Objectivity in Science, or Tom Kuhn Meets Tom Bayes. *Minnesota Studies in the Philosophy of Science*, **14**: 175 - 204.

Schindler, Samuel. 2018. *Theoretical Virtues in Science: Uncovering Reality Through Theory*. Cambridge: Cambridge University Press.

Schupbach, Jonah. 2011. Comparing Probabilistic Measures of Explanatory Power. *Philosophy of Science*, **78**: 813 - 829.

Tulodziecki, Dana. 2012. Epistemic Equivalence and Epistemic Incapacitation. *British Journal for the Philosophy of Science*, **63** (2): 313-328.
van Fraassen, Bas. 1980. *The Scientific Image*. Oxford: Oxford University Press.

Olfactory valence and theories of sensory pleasure

Błażej Skrzypulec

Olfactory valence and theories of sensory pleasure

In the case of major exteroceptive perceptual modalities, such as vision and audition, perceptual experiences are not usually associated with a strong valence-related component. While a visual stimulus might be repulsive and listening to music may be highly pleasurable, typical visual and auditory experiences seem to be generally neutral. The situation differs in the case of olfactory perception insofar as olfactory experiences commonly present odours as being pleasurable or unpleasant. In this respect, olfaction is similar to interoceptive modalities, like nociception, which are often associated with a negative or positive valence. In fact, scientists and philosophers investigating human olfaction often claim that odour perception is largely a perception of valence (Yeshurun and Sobel 2010). Valence is one of the main descriptors used in characterising olfactory stimuli (Khan et al. 2007) and olfactory mechanisms are closely associated with emotional responses (Keller 2016, 123–128; Soundry et al. 2011; Stevenson 2009).

Nevertheless, despite the significance of the hedonic aspect of olfactory perception and increased interest in olfaction among philosophers of perception (e.g., Aasen 2019; Batty 2010; Millar 2019; Skrzypulec 2019; Young 2019), little work undertakes a widespread evaluation of theories of sensory pleasure in the case of human olfaction. While identifying the proper approach to olfactory valence is in itself an interesting philosophical question, the issue also bears on a broader investigation of the philosophy of olfaction. Due to the significant role that valence plays in olfactory perception, it has been proposed that the role of olfaction is not to represent properties of stimuli, but rather to generate a valence-related response and motivate adaptive behaviour (Castro and Seeley 2014; Cooke and Myin 2011). It has been observed that similarity in olfactory valence is often unrelated to similarities between the molecular composition of stimuli and that other factors, such as the subject's beliefs or expectations, significantly influence hedonic aspects of olfactory experiences (see Barwich 2018, 2019; Keller 2016). Consequently, because valence is a crucial aspect of olfaction but is not strictly tied to the chemical properties of stimuli, one may doubt in a general program of treating olfactory experiences in a manner analogous to visual or auditory experiences – as representations which accurately or inaccurately present properties of entities in the environment. However, since there are influential representational theories of sensory pleasure (e.g., Bain 2013; Cutter and Tye 2011; Gray 2018; O'Sullivan and Schroer 2012) antirepresentational positions are only appealing if it is implausible to characterise the olfactory perception of valence in representational terms.

My aim is to investigate which of the major philosophical theories of sensory pleasure is the most plausible in the olfactory context. In analyzing this issue, I do not discuss the traditional, theoretical arguments provided for and against the most influential accounts. These traditional theories have well-recognised problems and proponents who have responded to the most major critiques (see Bain 2019; Bramble 2013; Heathwood 2007). Instead, I develop new arguments which directly bear on empirical knowledge concerning the olfactory perception of valence. In particular, I am interested in establishing whether olfactory pleasure and displeasure should be characterised in representational terms, since this issue is highly significant for the project of applying representational accounts of perception to the olfactory modality. I argue that how olfaction processes valence strongly suggests that olfactory valence should be, at least partially, understood per representational theories of sensory pleasure.

In particular, I introduce two conceptual distinctions which allow categorizing the theories of sensory pleasure in the context of olfaction. The first distinction differentiates between (a) World-Directed Theories according to which olfactory sensory pleasure and displeasure consist in having a mental state which can be 'successful' or 'unsuccessful' depending on the relation between the state's content and the world and (b) Not-World-Directed Theories according to which hedonic olfactory states do not have content allowing for such 'successfulness' or 'unsuccessfulness'. Relying on the behavioural results regarding olfactory valence perception (Herz and von Clef 2001; Wilson and Stevenson 2006), I argue in favour of World-Directed Theories as they are better suited for explaining the interactions between olfactory hedonic states and postperceptual propositional states like beliefs and expectations. The second distinction distinguishes two types of World-Directed Theories: (a) Representational Theories according to which olfactory valence consists in representing olfactory stimuli as having evaluative properties and (b) Satisfaction Theories which postulate that olfactory valence consists in having a state, such as a desire or a command, with content specifying what should, or should not, occur. By referring to the neuroscientific results on one- and two-dimensional coding of olfactory valence (Grabenhorst et al. 2007; Jin et al. 2015), I argue that the Representational Hypothesis is more plausible. It is so because the olfactory valence is likely to be, at least partially, coded one-dimensionally, and the Satisfaction Theories, but not Representational Theories, have difficulties in accounting for one-dimensional coding.

These results suggest that the way in which olfaction processes valence should be, at least partially, understood per representational theories of sensory pleasure. Consequently, I find no strong justification for positions that infer an antirepresentationalist conclusion from the fact that olfactory perception is largely valence-related.

Aasen, S. (2019). Spatial aspects of olfactory experience. Canadian Journal of Philosophy, 49(8), 1041-1061.

Barwich A.-S. (2018). Measuring the world. Olfaction as a process model of perception. In D. J. Nicholson, J. Dupre (Eds.), Everything flows: Towards a processual philosophy of biology (pp. 337-356). Oxford: Oxford University Press.

Bain, D. (2013). What makes pains unpleasant?. Philosophical Studies, 166, 69–89.

Bain, D. (2019). Why take painkillers?. Nous, 53(2), 463-490.

Barwich, A.-S. (2019). A critique of olfactory objects. Frontiers in Psychology, 10, 1337.

Batty, C. (2010). A representational account of olfactory experience. Canadian Journal of Philosophy, 40, 511–538.

Bramble, B. (2013). The distinctive feeling theory of pleasure. Philosophical Studies, 162, 201-217.

Castro, J. B., & Seeley, W. P. (2014). Olfaction, valuation, and action: reorienting perception. Frontiers in Psychology, doi:10.3389/fpsyg.2014.00299.

Cooke, E., & Myin, E. (2011). Is trilled smell possible? How the structure of olfaction determines the phenomenology of smell. Journal of Consciousness Studies, 18(11–12), 59–95.

Cutter B., & Tye, M. (2011). Tracking representationalism and the painfulness of pain. Philosophical Issues, 21(1), 90-109.

Grabenhorst, F., Rolls, E. T., Margot, C., da Silva, M. A. A. P., & Velazco, M. I. (2007). How pleasant and unpleasant stimuli combine in different brain regions: Odor mixtures. Journal of Neuroscience, 27(49), 13532-13540.

Gray, R. (2018). On the content and character of pain experience. Pacific Philosophical Quarterly, 100(1), 47-68.

Heathwood, C. (2007). The reduction of sensory pleasure to desire. Philosophical Studies, 133, 23-44.

Herz, R. S., & von Clef, J. (2001). The influence of verbal labeling on the perception of odors: Evidence for olfactory illusions?. Perception, 30, 381-391.

Jin, J., Zelano, C., Gottfried, J. A., & Mohanty, A. (2015). Human amygdala represents the complete spectrum of subjective valence. The Journal of Neuroscience, 35(45), 15145–15156.

Keller, A. (2016). Philosophy of olfactory perception. New York: Palgrave Macmillian.

Khan, R. M., Luk, C.-H., Flinker, A., Aggarwal, A., Lapid, H., Haddad, R., & Sobel, N. (2007). Predicting odor pleasantness from odorant structure: Pleasantness as a reflection of the physical world. The Journal of Neuroscience, 27(37), 10015–10023.

Millar, B. (2019). Smelling objects. Synthese, 196, 4279-4303.

O'Sullivan, B., & Schroer, R. (2012). Painful reasons: representationalism as a theory of pain. The Philosophical Quarterly, 62(249), 737-758.

Skrzypulec, B. (2019). The nonclassical mereology of olfactory experiences. Synthese, https://doi.org/10.1007/s11229-018-02072-x.

Soudry, Y., Lemogne, C., Malinvaud, D., Consoli, S.-M., & Bonfils, P. (2011). Olfactory system and emotion: Common substrates. European Annals of Otorhinolaryngology, 128, 18—23.

Stevenson, J. (2009). Phenomenal and access consciousness in olfaction. Consciousness and Cognition, 18, 1004–1017.

Wilson, D. A., & Stevenson, R. J. (2006). Learning to smell: olfactory perception from neurobiology to behavior. Baltimore: The John Hopkins University Press.

Yeshurun, Y., & Sobel, N. (2010). An odor is not worth a thousand words: from multidimensional odors to unidimensional odor objects. Annual Review of Psychology, 61, 219-241.

Young, B. D. (2019). Smelling molecular structure. In D, Shottenkirk, S, Gouveia, J. Curado (Eds.), Perception, cognition, and aesthetics (pp. 64–84). New York: Routledge Press.

The Ontological and Epistemological Implications of Using Bottom-Up Statistical Analysis to Establish Dimensional Systems of Psychopathology: A Preliminary Roadmap

Helo Liis Soodla and Kirsti Akkermann

Philosophy of Cognitive and Behavioral Sciences

Title: The Ontological and Epistemological Implications of Using Bottom-Up Statistical Analysis to Establish Dimensional Systems of Psychopathology: A Preliminary Roadmap

Keywords: philosophy of psychiatry, psychiatric taxonomy, philosophy of statistics

Short abstract (81 words)

We argue that psychiatric nosology moving from categorical classification to statistically rigorous dimensional classification requires two methodological shifts. Firstly, conceptual analysis of classification entities should move beyond the realist-non-realist and essentialist-non-essentialist dichotomies to account for ambivalence created by use of bottom-up data-driven classification. Secondly, consideration of both traditional issues in nosology research and additional statistical constraints meaningfully impacts psychiatric research methodology. We exemplify these claims by providing a preliminary philosophical account of latent profile analysis as means for establishing dimensional nosologies.

Extended abstract (987 words)

The concept of a "disorder" has been widely investigated within philosophy of psychiatry, with discussion mostly focusing on the ontology of categorical disorders as defined by the International Classification of Diseases and the Diagnostic and Statistical Manual of Mental Disorders (Kendler et al., 2011). Such classification systems have two main features:

(1) individuals are classified as presenting or not presenting with a specific disorder;

(2) the necessary symptoms that are altogether sufficient for making a diagnosis have been determined by expert consensus, rather than empirical methods (Krueger & Piasecki, 2002).

As such, categorical taxonomies have been criticized for their insufficient explanatory power and failure to account for varied disorder phenomenology, prompting the creation of alternative, dimensional nosologies (Widiger & Gore, 2014). Such approaches have two main features:

(1) individuals are characterized by traits that vary on a continuum, rather than categorical ascription of diagnosis;

(2) the relevant traits are selected and combined into profiles via bottom-up empirical methods. (Krueger & Piasecki, 2002).

We argue that such a paradigm change requires recalibration on part of philosophy of science and that the ontological and epistemological implications arising from changing classification procedures meaningfully impact psychiatric research. We exemplify these claims by presenting a preliminary analysis of dimensional profile-based classification systems. Firstly, we outline the main strands of philosophical research on categorical disorders. Secondly, we highlight the necessity of amending the analytical toolbox to account for dimensional classification systems. Finally, we suggest a putative roadmap for future psychiatric research.

We begin by delineating central dichotomies in the discussion on psychiatric nosologies. In describing categorical disorders as entities, authors have either been arguing for a realist or a non-realist position and either an essentialist or a non-essentialist position (Zachar & Kendler, 2017). Robustly put, to be a realist in terms of categorical disorders is to claim the perceiver-independent existence of "something out there" (Kendler et al., 2011). Non-realism can take on many forms: disorders could be socially constructed entities or pragmatic/instrumental constructs the metaphysics of which are simply not relevant (Kendler et al., 2011). Essentialist accounts have been equated to natural kind theories according to which all instances of the entity have an infinite number of characteristics in common that together make up the essence of the disorder (Werkhoven, 2021). Non-essentialist positions can align with non-realists but need not: psychiatric disorders have also been considered property clusters joined together by causal mechanisms but not defined by common essences (Boyd, 1991).

We propose that such distinctions are useful, yet insufficient for describing the questions relevant for researching dimensional classification systems and integrating them into practice. Such classification derives from statistical analysis: e.g. factor analytic methods, latent profile analysis, clustering or structural equation modelling. We argue that reliance on specific statistic models puts constraints on how resulting solutions can be interpreted.

Consider the example of latent profiling. Latent variable models presume the existence of underlying unmeasurable traits that cause variance in measurable traits and this latent variable membership explains any shared variance among the indicators (Nylund-Gybson & Choi, 2018). To perform a latent profile analysis, a set of continuous indicator variable data – the traits that are expected to contribute to the creation of dimensional psychopathology profiles – is subjected to an algorithm with pre-determined number of profiles to be extracted (Tein et al., 2013). Latent profile models are assessed via goodness-of-fit indices that reflect how well a model represents the data (Tein et al., 2013).

Resulting profiles could be analysed as entities and in terms of whether they align with realist and essentialist predictions. However, reliance on statistical assumptions creates additional ambiguity. Firstly, a latent profile analysis always results in extracted profiles (if there are no serious problems with non-convergence or model underspecification). However, we do not seem to believe there to be an infinite number of psychopathology entities. Secondly, the researcher determines the number of profiles to be extracted, yet we do not intuitively believe the statistical procedure of extracting $n + 1$ classes to create an additional entity in the world. Thirdly, goodness of fit indices can result in conflicting fit judgments (across, say n, $n + 1$, $n + 2$ profile models), however, there is no non-empirical benchmark to calibrate the fit indices. Borsboom (2017) postulates an internal tension within such measurement models: representation in statistical notation and explanation in conceptual analyses do not seem to align. If we took dimensional profiles to be merely statistical concepts with no explanatory component, much of contemporary research would become nonsensical (analyses run on different samples and with different analysis parameters would all result in new statistical "entities"); if we did not postulate the representational component, our conclusions would be unmerited (the charm of the rigorous "bottom-up" approach lost).

Rather than offering a one-size-fits-all solution to this conundrum, we propose ways in which such considerations should be addressed in practice. In conducting research, we stress the importance of bridging the traditions of diagnostic classification and inferential statistics to produce a set of questions to inform methodological choices and their interpretation. Firstly, the kinds of presumptions made by employing a certain analytic strategy should be made explicit. For example, it seems to be implicitly assumed that model fit to data reflects the extent to which a theory based on said data "carves nature at its joints", yet such an assumption is not necessarily merited. Additionally, analytic solutions are interpreted as being theoretically meaningful, even if the statistical analysis is theoretically agnostic and produces latent profiles based on simple mathematical algorithms. Secondly, the objective of classification should be made explicit: is the goal to produce a heuristic for practitioners, a descriptive distinction for the medical system, a nosology aiming to achieve theoretical coherence or a sophisticated statistical model. Thirdly, to inform interpretation, it should be assessed, to what extent the assumptions made by statistical methods align with set objectives. Finally, to address these three questions, methodological choices, and the rationale behind them should be reported on meticulously.

References:

Borsboom, D. (2017). Representation and explanation in psychometric modeling. In K. S. Kendler & J. Parnas (Eds.), *Philosophical issues in psychiatry IV: Classification of psychiatric illness* (pp. 45–50). Oxford University Press.

Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *61*(1/2), 127–148.

Kendler, K. S., Zachar, P., & Craver, C. (2011). What kinds of things are psychiatric disorders? *Psychological Medicine*, *41*(6), 1143–1150. https://doi.org/10.1017/S0033291710001844

Krueger, R. F., & Piasecki, T. M. (2002). Toward a dimensional and psychometrically-informed approach to conceptualizing psychopathology. *Behaviour Research and Therapy*, *40*(5), 485–499. https://doi.org/10.1016/s0005-7967(02)00016-5

Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science,* *4*(4), 440–461. http://dx.doi.org/10.1037/tps0000176

Tein, J. Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(4), 640–657. https://doi.org/10.1080/10705511.2013.824781

Werkhoven, S. (2021). Natural kinds of mental disorder. *Synthese, 199*(3), 10135–10165. https://doi.org/10.1007/s11229-021-03239-9

Widiger, T. A., & Gore, W. L. (2014). Dimensional versus categorical models of psychopathology. *The Encyclopedia of Clinical Psychology*, 1–12. https://doi.org/10.1002/9781118625392.wbecp108

Zachar, P., & Kendler, K. S. (2017). The philosophy of nosology. *Annual Review of Clinical Psychology*, *13*, 49–71. https://doi.org/10.1146/annurev-clinpsy-032816-045020

Objectivity in Practice: Disenchanting AI

Mark Theunissen and Jacob Browning

In a recent article, Inkeri Koskinen (2021) argues that claiming something is objective is a matter of endorsing it because it minimizes the risk of epistemic bias. But she notes this presents us with a dilemma when dealing with AI systems: either treat ML systems simply as tools or treat them as their own agents. On the one hand, treating AI as a tool means we focus on whether they are objective—whether they should be endorsed because they help reduce human epistemic bias. This is an attractive option since these systems are tools and need to be integrated into the professional's practice if they are to be useful. On the other hand, treating AI as a cognitive agent seems appropriate as well, since they may have machine-specific biases. If they are cognitive agents like humans, but with different epistemic weaknesses, then we should not rely on them simply as tools.

We argue that the dilemma Koskinen identifies is better understood as a spectrum, where some ML systems are better treated as tools and others as epistemic agents with specific notion of objectivity are at stake. Specifically, in the context of healthcare, when such ML systems are more akin to other medical tools, it is essential to evaluate them cooperatively to ensure they make the medical professionals using them more reliable. By contrast, if the system is functioning as a second opinion, they should be evaluated competitively because evidence shows professionals—such as doctors—are influenced by their outputs (Gaube et al. 2021). Treating them as falling along the spectrum highlights the different epistemic risks various systems pose, as well as suggesting ways to address these risks.

The spectrum approach also sheds light on an important related problem for ML systems: explainability. There is a lively debate about whether we need post hoc explanations at all or whether high accuracy in a system is sufficient for justification. But we contend the issue is not whether we need a post facto explanation at all, but when an explanation is called for—and when accuracy is sufficient. As we will show, highly accurate systems falling on the tool side of the spectrum gain few epistemic benefits from explainability; treating them as objective is sufficient. By contrast, when the ML system is better understood as an epistemic agent, post hoc explanations are often essential for reducing the opacity of the decision provided and making it possible for fellow (human) agents to effectively work with and make use of the system.

Bibliography

Gaube, S., Suresh, H., Raue, M., Merritt, A., Berkowitz, S. J., Lermer, E., ... & Ghassemi, M. (2021). Do as AI say: susceptibility in deployment of clinical decision-aids. NPJ digital medicine, 4(31), 1-8. https://doi.org/10.1038/s41746-021-00385-9

Koskinen, I. (2020). Defending a risk account of scientific objectivity. The British Journal for the Philosophy of Science, 71(4), 1187-1207.

Model Transfer and Universal Patterns - Lessons From the Yule Process

Sebastiaan Tieleman

Model Transfer and Universal Patterns - Lessons From the Yule Process

Abstract

Model transfer refers to the observation that particular model structures are used across multiple distinct scientific domains. This paper puts forward an account to explain the inter-domain transfer of model structures. Central in the account is the role of validation criteria in determining whether a model is considered to be useful by practitioners. Validation criteria are points of reference to which model correctness for a particular purpose is assessed. I argue that validation criteria can be categorized as being mathematical, theoretical or phenomenological in nature. Model transfer is explained by overlap in validation criteria between scientific domains. Particular emphasis is placed on overlap between phenomenological criteria. Overlap in phenomenological criteria can be explained through the notion of universal patterns. Universal patterns are abstract structures that can be made to refer to multiple distinct phenomena when coupled with phenomena-specific empirical content. I present the case study of the Yule Process, in which universal patterns play a crucial role in explaining model transfer. This paper provides an account of model transfer that stays close to modelling practice and expands existing accounts by introducing the notion of universal patterns.

1 Introduction

An observation in the use of models in science is that particular model structures are used across multiple distinct scientific domains. To clarify, model here refers to what has been described as a model-type (Van Fraassen et al., 1980); a model in which parameter values may remain unspecified. In this context, model structure refers to the abstract structure of a model-type, meaning that, in the case of mathematical models, the variables and parameters of the structure do not refer to anything that can be observed empirically. The observation of a model structure that is imported into a new domain can be labelled as inter-domain model transfer. As an example, the growth process of firms is modelled using the same mathematical structure as the Yule process, which is a model originally developed in evolutionary biology (Simon, 1955). Such observations contrast a view of science in which various scientific domains operate in isolation, each using a domain specific methodology.

Instead, we can view science as being organized through a particular set of methods, including certain formal structures used in modelling (Humphreys, 2004). This observation is puzzling however, when we consider that models are, generally speaking, constructed for a domain-specific purpose: answering a question (Boumans, 2006). Such questions often concern phenomena. For example, how do firms grow in size over time (Simon & Bonini, 1958)? Questions about phenomena are inherently domain specific; they ask about a growth process of, in this example, a specific economic entity, firms. The ability of a model to answer this question is usually built into the structure of the model (Boumans, 1999), by shaping the model structure in such a way that it fulfils relevant validation criteria. Perhaps one would expect that a model structure shaped by validation criteria that are deemed relevant for a domain specific purpose would always produce a domain specific model structure, but for some particular model structures this is not the case.

The main question that this paper will seek to answer is what explains the inter-domain transfer of some model structures? Implicit in the observation that some model structures are transferred across multiple domains is that these model structures are somehow considered to be useful across the domains to which they are applied. Another way of putting this question is, therefore, what makes a model structure useful in the domain it was constructed for, as well as the domain it is transferred to? In order to answer this question, this paper will introduce a novel framework of model transfer. The foundation for this framework is the model construction account by Boumans (1999). It entails that models are constructed such that they meet various validation criteria. Validation is defined here as the broad assessment of model correctness in relation to its purpose. Validation criteria are points of reference to which model correctness for a particular purpose is assessed. For example, we could assess whether a model is in line with relevant theory or we could assess whether the model is able to reproduce certain facts about phenomena. It is the fulfilment of such validation criteria that determines whether a model is considered to be useful. Given this account I will show that inter-domain model transfer can be explained by overlap between validation criteria across domains. Special attention will be paid to overlap between so-called phenomenological validation criteria. To explain how this overlap can occur, I will introduce the notion of universal patterns. Universal patterns are abstract structures that, when coupled with empirical content, can be made to apply to multiple distinct domains. Empirical content refers to the information that relates an abstract structure to objects that can be observed empirically (Humphreys, 2019). In order to illustrate my analysis I will discuss a case study of model transfer. The study concerns the Yule Process, a model that was first developed in evolutionary biology (Yule, 1925), and later transferred to various other systems including the growth of firms (Simon, 1955).

In the existing literature we can distinguish three main accounts that seek to explain model transfer (Knuuttila & Loettgers, 2020),analogues (Hesse, 1966) which attributes model transfer to similarity relationships between phenomena, formal templates (Humphreys, 2019), which attributes model transfer mainly to overlap in construction assumptions and model templates (Knuuttila & Loettgers, 2016), which attributes model transfer to overlap in conceptual features. What each of these accounts embed is a notion of inter-domain model usefulness. They point to particular aspects of model structures that allow scientists to re-use these structures across distinct domains. I will argue, however, that although valuable, these accounts do not give a complete enough description of what it is that makes a model considered to be useful in practice.

Looking at models as analogies is an account discussed in a.o. Hesse (1966). Within this account, models derive utility from the similarity relations they have with the phenomenon of interest. Hesse (1966) distinguishes positive, negative and neutral analogies. In the context of models, positive analogies are the aspects of the phenomenon of interest and the model mechanisms that overlap. Negative analogies are the aspects that do not overlap. Neutral analogies are the aspects for which this overlap is yet to be determined, and are thus what makes the model potentially useful to learn about the phenomenon of interest. In order for the structure of a model be useful, it must thus be a positive analogy of the phenomenon of interest in that particular domain to some degree. In the case of model transfer, this implies that features of the model structure are a positive analogy in both the original and new domain. This is likely to be the case when there is a similarity relation between targeted phenomena of the different domains. If we consider a model of genera growth in biological evolution, the structure of which is also used as a model for city growth, for example Simon and Bonini (1958), it is likely that there are certain features in the model that serve as analogies to genera growth in biological evolution as well as firm growth. Importantly, such features cannot be domain specific, and are thus to some extent abstract. As we will see in the case study of this paper, one of these features is proportional growth, which can serve as an analogy to how both genera and firms grow. What is transferred according to this account, is thus an analogy that applies to multiple domains. This still leaves open, however, why it is that certain abstract features can serve as positive analogies in multiple domains. Furthermore, as also noted in Humphreys (2019), such analogies can often be made to fit in a domain opportunistically. Just looking at model transfer in the context of analogies may thus not always yield a satisfactory account of model transfer.

A different view comes from Humphreys (2004) in which the idea of a computational template is put forward. A computational template is a computational structure that can be adjusted to be used as a model in distinct domains. The utility of using this computational template and the explanation as to why some model structures become templates are favourable analytical-tractability properties. The template should also be flexible; it should be open to adjustments, such that it can be made to fit various distinct domains. This view of model transfer, however, was originally put forward, to be applicable to computational models. More recently, we have seen an extension of this account in Humphreys (2019). This view regards that what is being transferred a so-called formal template. In this account, the usefulness of a model structure is essentially determined by the correctness of a model's construction assumptions. Model transfer, in this account, is therefore enabled by the correctness of the construction assumptions in the original and new domain on a more abstract formal level. If a construction assumption is a linear relationship between two variables then this assumption should hold in both domains. That what is transferred in essence is thus not an analogy, but a "correct" formal structure with favourable formal properties. Knuuttila and Loettgers (2020) state, however, that just considering formal properties is not a complete explanation because it does not explain why some model structures are transferred between domains widely and others are not. Many model structures that are successfully used within a particular domain will have favourable formal properties such as analytical tractability. Only few, however, are transferred across domains.

Another important addition to the model transfer literature is (Knuuttila & Loettgers, 2016), in which the concept of a model template is introduced. This is a template with favourable formal properties coupled with general conceptual features. These conceptual features suggest how to theorise about the phenomenon described by the model. This implies that model transfer is enabled when the conceptual features embedded in the template are deemed useful tools for theorising in both the original and new domain. Examples of such conceptual features are given in Knuuttila and Loettgers (2020) include phase transitions and local interactions. The account of model templates points to a particular source of model usefulness that allows us to explain some instances of model transfer. The account, however, is, in my view, most applicable to the methods and conceptual notions present in complexity science and, therefore, limited in its scope of application.

The essential difference between the account of model transfer put forward in this paper is that it is does not rely on a particular epistemological account of model usefulness. Instead, rather than explaining what makes a model structure useful, I will take a more empirical approach and look at what makes a model structure *considered* to be useful in observed scientific practice. This approach in my view, results in an account of model transfer that is a closer match to scientific practice and, therefore, covers a wider range of model-transfer cases. It also does not rely on a particular epistemological view of model usefulness. Furthermore, it highlights an enabling factor of model transfer that is not explicitly present in the accounts of model transfer discussed, namely universal patterns. The account presented here is also general in the sense that it subsumes the existing accounts of model transfer here to some extent.

The accounts described above rely on a particular epistemological account on model usefulness. Such accounts, even though valuable, risk being an incomplete match to observed scientific practice. In order for an account of model transfer to be a correct explanation of scientific practice, a more empirical approach is required. Rather than explaining what makes a model structure useful, I will look at what makes a model structure *considered* to be useful in observed scientific practice. This approach, in my view, results in an account of model transfer that does not rely on a particular epistemological view on model usefulness and, therefore matches more closely with scientific practice.

To specify the aforementioned criteria of model usefulness, I build on the literature on model validation, which I have defined as the assessment of a model's correctness relative to its purpose. The benchmarks in the validation process are validation criteria. To this regard, Boumans (1999) shows that the ability of the model to fulfil such criteria is built into the model, and is thus central in shaping the model structure. To assess whether the model is able to fulfil these validation criteria to a satisfactory degree, the model is subjected to various validation tests (Senge & Forrester, 1980). Which validation tests are deemed relevant, differs given the purpose of the model (Barlas, 1996). Looking at model transfer from the point of view of validation, model transfer is enabled by satisfactory validation in the original and the new domain, which, in turn, is enabled by overlapping validation criteria. In this paper, I will argue that empirical validation may play a key role in the transfer process, meaning the assessment of whether the model is able to reproduce relevant facts about phenomena. In such cases, the model structure that is transferred must be able to reproduce facts about phenomena in the original as well as the new domain. Empirical validation as a mechanism of model transfer is supported by the notion of universal patterns. Universal patterns help us understand why certain model structures are transferred so widely.

An account of model transfer that also starts from scientific practice can be found in Donhauser (2020). It contrasts two opposing viewpoints regarding the ability of scientists within a particular domain to import knowledge from other scientific domains. Incommensurability states that epistemology is domain specific to such a large degree, that knowledge transfer between domains is impossible. On the other end, there is the notion of voluntarism, which states that scientists can "choose" a particular epistemological stance as long as certain general conditions are met. Donhauser (2020) argues that incommensurability is not able to explain model transfer while voluntarism does. As we will see, the idea put forward in this paper fits neither of these epistemological viewpoints perfectly. Instead, I will argue that models are likely to be transferred when there is overlap in the criteria used to assess model usefulness. The criteria scientists use do not necessarily have to be the result of voluntary decisions under general conditions, but may also be a function of particular paradigms. As is argued in Humphreys (2004), a paradigmatic organisation of science is not necessarily domain specific. Rather, certain methodological strategies span multiple distinct domains.

The reader may associate the notion of model validity with the notion robustness, or, more specifically, with the notion of model robustness such as put forward in Lloyd (2015). Model robustness refers to a degree of insensitivity of a model's ability to reproduce facts about phenomena, to changes in various assumptions and/or parameter values of the model. Inter-domain model transfer could be seen as robustness with respect to changes in the empirical content of a model structure. If we change the empirical content of a model structure (transfer a model structure to a new domain), the model is still able to reproduce relevant facts about phenomena. Generally speaking, however, robustness refers to a property of model structures that reproduce facts about phenomena with the same empirical content. Therefore, to avoid confusion, I will not engage explicitly with the notion of model robustness in relation to model transfer. Assessment of model robustness, as it is generally understood, however, may be subsumed in the more general empirical validation process when relevant. Often the assessment of model robustness may come in the form of sensitivity analysis; altering parameter values and/or model assumption and assessing how this affects model output.

2 Framework: Validation Criteria and Model Transfer

Central in what I argue in this paper is that satisfactory model construction requires fulfilment of certain validation criteria (Boumans, 1999). The model structure is, therefore, shaped by its validation criteria. This implies that the model can only be reused in a new domain when it can be validated within this new domain. Given the account of model construction that I will present here, this is the case if and only if there is overlap in the validation criteria in both the original and the new domain. Let us now take a closer look at the account in Boumans (1999) to understand, first, what validation criteria consist of more specifically and second, how they are part of the construction process.

The validation criteria are determined in relation to the purpose of the model. There are multiple ways in which we could classify different types of validation criteria. For the purposes of our framework, I distinguish between theoretical, mathematical or phenomenological criteria, which stays close to the types of criteria mentioned in Boumans (1999). Theoretical criteria include questions like: is the answer provided by the model, to some extent, in line with what we would expect from theory X? Given the law of supply of demand in economics for example, a criterion could be that the model incorporates a negative relationship between price and demand (ceteris paribus). Mathematical criteria may include criteria of analytical tractability, the model must not be so complex that it does not enhance understanding. Finally, phenomenological criteria can come in the form of empirical validation; is the model able to reproduce fact Y? Importantly, of course, all of these criteria must be relevant to the purpose of the model (Boumans, 2009). Relevance for the three types of justification criteria includes the following: First, the theoretical criteria should involve theories that have implications for the question at hand. Second, the strictness of analytical tractability criteria depends on whether the model's purpose is to provide understanding of certain mechanisms. If a model's purpose is solely to predict, for example, strict analytical tractability criteria are not relevant. Third, the facts to reproduce should be relevant to the explanation the model provides. If the purpose of the model is to provide an explanation of a particular phenomenon, the facts to be reproduced by the model are usually facts about that particular phenomenon. To illustrate, a model constructed to explain the business cycle in economics is usually required to be able to reproduce the empirically observed business cycle.

Models go through a process of construction. They are not just discovered, and are not a trivial extension of theory. The question is, however, whether this construction process is independent from the above described validation process. In a more traditional view, these processes are considered as independent, which roughly means that the validation process starts after the model is constructed. If the model fails to pass the validation criteria, the model is to be discarded. As shown through case studies in Boumans (1999), the problem with this traditional view is that it is not in line with actual scientific practice. Given that the validation criteria are given by the question the model is constructed to answer, they are known during the construction process, and play an important role in the construction process. Models are constructed in such a way that the model meets the criteria. When the model does not meet the criteria a "back and forth" process starts in which the model is tweaked and altered until the criteria are met to a sufficient degree. The ability of the model to meet its validation criteria is thus built into the structure of the model. This concerns all three theoretical, mathematical, and phenomenological criteria. The case studied in Boumans (1999) for example, concerns how (in addition to theoretical and mathematical criteria) a microfounded business cycle model is constructed to reproduce the Phillips-Curve (the negative relationship between inflation and unemployment), which is a phenomenological criterion.

An additional element that may be considered, is that the ability of a model to fulfil one validation criterion is often not independent from the fulfilment of the other validation criteria. This implies that model construction, in practice, often comes down to a balancing act between the various relevant validation criteria. As an example, there may be tension between the fulfilment of theoretical and mathematical criteria. Theoretical notions may be complex to such a degree that their incorporation into a model structure would cause the model to become analytically unsolvable, or the model could become so complex that it is unintelligible. As we will see in the case study presented later, the balancing of theoretical and mathematical criteria was an explicit issue in Yule (1925). In the same way, theoretical and phenomenological criteria may be at odds. The incorporation of certain theoretical notions into a model structure may imply that the model output is not in line with certain facts about phenomena. In some instances, the modeller has to prioritize certain validation criteria. As I will discuss in more depth in the case study later in this paper, for example, the starting point for the model presented in Simon and Bonini (1958) was a dissatisfaction with microeconomic theory because of its inability to reproduce the observed distribution of firm size. Of course any balancing or prioritisation of validation criteria is again a function of the purpose of the model.

A further complicating factor may be that some validation criteria in practice cannot be identified as being purely theoretical, mathematical or phenomenological. For example, the theoretical notions that underlay what we could recognise as theoretical validation criteria, may themselves be partially based on empirical evidence. In addition, in models in physics in particular, theoretical notions are sometimes tied to particular mathematical formulations. Being able to express a theoretical notion with mathematical elegance is sometimes seen as support for that theoretical notion. Often, however, as we will also see in the case study, we are able to classify a criterion as being primarily theoretical, mathematical or phenomenological.

This account of model construction applies to model that are constructed from the ground up as well as models that re-use existing model structures. Models constructed by recycling existing model structures are also subject to the various types of criteria outlined above. For model structures to be acceptable in both the original and new domain, there must thus be overlap in the validation criteria. In the framework presented here, overlap in validation criteria are what enables model transfer across distinct domains. To clarify, we can look at the three main types of validation criteria distinguished before. In the case of theoretical criteria there may be overlap if the core idea of the theory is sufficiently abstract. We can think of certain concepts from evolutionary theory that are considered useful in biology but also in some sub-fields of economics (Dosi & Nelson, 1994). In the case of mathematical criteria, it is not hard to see that, for example, analytical-tractability criteria may apply across distinct domains. Finally, in the case of overlap in phenomenological criteria, we can think of requiring models to reproduce the same type of empirically observed distribution in the original and new domain. The account of a model template by Knuuttila and Loettgers (2016) can be seen as a vehicle for the fulfilment of theoretical and mathematical criteria. I argue that this account risks being incomplete in cases where it is overlap between phenomenological criteria enables model transfer. One may wonder how it is that certain facts about phenomena will be the same across distinct domains. In the next section, I will provide an explanation for the occurrence of overlap in phenomenological criteria.

We may posit that fulfilling these validation criteria shows some similarity relationship between the model structure and the real world structure and, in the case of model transfer, is thus evidence of a similarity relation between the targeted real world structure of the original and the new model, which is also implied by an account that looks at models as analogies such as Hesse (1966). This depends, however, on the relationship between the fulfilment of validation criteria and the representational value of the model. I argue that it is not useful to consider this relationship for the purpose of this paper. First, this relationship is complex and uncertain and depends to a large extent on whether one holds a realist or more instrumentalist stance towards scientific models (Gatti, Fagiolo, Gallegati, Richiardi, & Russo, 2018). Second, as is also shown in Barlas (1996) it depends on the purpose of the model. For so-called, black-box models, for example, the sole purpose of the model is to give correct predictions which implies that the representational value of the model mechanisms are not a relevant criterion of assessment. Not directly engaging with the relationship between validation criteria and the representational value of the model is thus more epistemologically neutral and covers a wider range of model-types.

3 Universal Patterns

I have stated that overlap in phenomenological criteria should be taken into account in order to come to a more complete account of model transfer. The question that remains to be answered is when is this the case? Empirical validation tests generally consist of assessing whether the model is able to reproduce relevant facts about phenomena. Overlap of phenomenological criteria implies, therefore, that there is somehow overlap in features of these facts about phenomena. This may seem unlikely given that facts about phenomena are associated with something that is tied to empirical content, namely a phenomenon. The distribution of firm size, is about a specific domain, firms. Abstract features of such facts, however, may very well appear across multiple distinct domains. These features are what I will label as universal patterns. As we will see, the distribution of firm size follows a particular power law, the Yule Distribution, which is a feature of many observed distributions in distinct domains (Simon, 1955).

Let me first elaborate what I mean exactly by a universal pattern. A pattern can be thought of as an abstract structure. It is abstract because, by itself, the pattern does not have any empirical content, meaning that it neither empirically true or false (Humphreys, 2019). It simply does not refer to any object that can be observed empirically. It is a structure because we perceive it as something structured as opposed to being unstructured. Typical structures would be geometric shapes, like circles, curves, cycles and spirals, or it may also be structured in the sense that they can be described by a particular mathematical form. As an example of an abstract structure, we can think of patterns used in knitting; even though the patterns by themselves do not refer to anything empirical, we still recognize them as having a structure. Patterns can be made to refer to specific facts about phenomena by coupling them with specific empirical content. Empirical content, in this sense, refers to the information that relates the abstract structure to the empirically observable facts about phenomena. When the Yule Distribution is used as the distribution of genera size, for example, it is coupled with information that gives particular meaning to the shape. A point on the line that is higher than another point on the line, means that the higher point represents a genus that is larger in terms of species. Note that there are four relevant concepts within this description: the pattern, the empirical content, the fact about the phenomenon and the phenomenon itself. Patterns can be made to match a fact about a phenomenon by coupling it with empirical content. A pattern is a universal pattern if and only if it can be made to refer to facts about phenomena in multiple domains by changing just the empirical content that the pattern is coupled with. In Figure 1, we can see a schematic overview to clarify the relationships between concepts. A single universal pattern can be made to apply both to fact about phenomenon A and B by coupling it with empirical content A and B respectively.

The notion of universal pattern put forward here is induced from the observation that certain patterns are observed and used in scientific practice in varying domains. Most straightforwardly, we can think of the Gaussian or normal distribution, which is observed across widely varying domains such as the human height or the weight of loaves of bread (Lyon, 2014). Another example are certain power distributions such as Zipf's law (Corominas-Murtra & Solé, 2010) or the Yule distribution (Simon, 1955) which are observed in the distribution of city size and the distribution of words in a piece of literature. Universal patterns are not limited to distributions however. We can think of particular oscillation patterns for example, which are observed in (among many other domains) ecology and economics (Gandolfo, 2008).

Let me now relate the notion of universal patterns more explicitly to what we have established in the previous sections. In order for a model to be transferred across domains it must be considered useful by the practitioners in both the original and the new domain. This usefulness is considered by assessing whether the model is able to meet certain validation criteria. These validation criteria are built into the structure of the model meaning that the model structure is shaped by the criteria. For a model to be useful in a domain

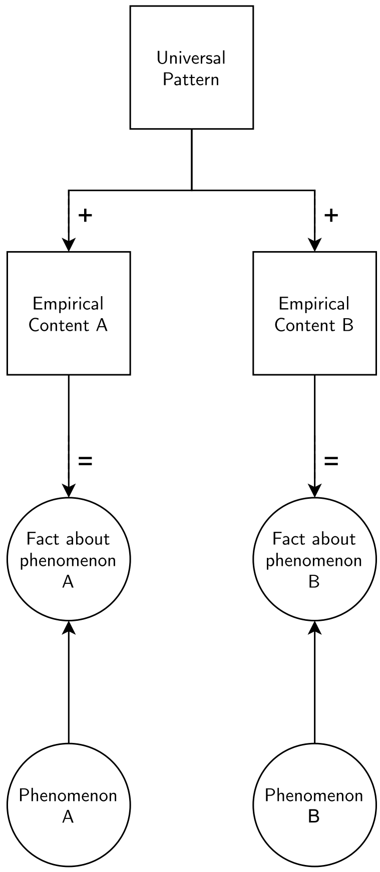


Fig. 1: Universal patterns and facts about phenomena

different from the one it was originally constructed for, the validation criteria should overlap. When phenomenological criteria have played an important role in shaping the structure of the original model, it is these criteria that should overlap in the new domain in order for the model structure to be transferred. This is the case when the phenomenological criteria embed a universal pattern.

The broad view is thus that in most modelling exercises there is a desire to latch the model onto the empirically observable world in some way. The observations we make, and the facts about phenomena we distil from them, are sometimes structured in specific ways. In such cases, models that are constructed to latch onto phenomena are likely to have a structure that is specific to that observed phenomenon. Devoid of any empirical content, such a fact about a phenomenon does not represent a universal pattern. In other instances, however, the facts about phenomena that we distil from our observations are structured in general ways. That is, they embed a pattern that can be made to refer to distinct facts about phenomena, a universal pattern. We are thus confronted with a world in which we do observe both specificity as well generality. Where we observe specific patterns, there likely are methodological borders. Where we observe universal patterns there likely are methodological transfers. This view contributes to an explanation for the observation that some particular model structures are transferred and not others.

The notion of universal patterns that I have presented here, is related to, but different from the existing concept of universality. The field that has discussed this notion of universality most explicitly is that of statistical mechanics. In statistical mechanics, universality concerns similarities in the behaviours of diverse systems (Batterman, 2000). Another way in which this is sometimes formulated is that the system level behaviour is independent to elements of the microscopic structure system (Batterman, 2000). If this is the case, it may imply that systems constituted of different objects still show similar behaviour. An example often used is when a magnet is heated to a certain critical temperature, it will lose its magnetism (phase transition). The path between these two states as a function of temperature (coexistence curve) is described by a power function with a critical exponent close to 1/3 (Batterman, 2000). The same functional form and critical exponent is also observed in phase transitions between the fluid and vapour states of matter like that of water. Clearly, the microscopic structure of water and magnets is different. Still, some properties at a system level are strikingly similar. The same notion of universality has also been applied to systems outside of chemistry and physics, such as agent-based systems (Parunak, Brueckner, & Savit, 2004) and biological systems (Batterman & Rice, 2014). The power function with a critical exponent close the 1/3 falls within the account of a universal pattern presented here. State transitions in matter and transitions in magnetism are facts about phenomena with distinct empirical content, but nonetheless express a similar pattern. The account of universal patterns that I have presented, however, does not make any statements about the relation between the observed pattern and the system it is generated by. In the statistical mechanisms notion, universality is a property of a system the behaviour of which comes in the form of widely observed patterns. This prepossess, however, that what is observed, is strictly tied to the system it is generated by. As I will discuss in the next paragraph, this limits the ways in which we can explain why we observe universal patterns, in a way that is not necessary within the context of model transfer.

Why we observe universal patterns is a fundamental question that requires a full investigation on its own and is thus beyond the scope of the main question of this paper. Generally, however we can distinguish between two types of explanations. One explanation comes from the same statistical mechanics notion discussed in the previous paragraph, and is discussed in a.o. Batterman and Rice (2014). It states that systems, even though being distinct in certain ways, still share abstract fundamental features, such as locality, conservation and symmetry. Such features provide an attractive fixed point such that systems that are different in some aspects, but share these fundamental features, converge to having the same properties, in the form of universal patterns. This explanation is related to the notion of a causal core as discussed in (Lloyd, 2015). The causal core consist of those features that are responsible for generating particular output, and are robust against changes that are outside this causal core. For physical systems, this explanation may seem credible, as stated before, however, universal patterns are also observed in diverse social phenomena (Simon, 1955). It might be less clear that such patterns are also the result of abstract fundamental features in the systems that they are generated by. According to some, however, this is the case. Mandelbrot and Hudson (2007), for example, applies they theory of fractals (Mandelbrot, 1982) as an explanation for the distribution of price changes on stock markets. Fractals are seen by some as a fundamental self-organizing principle of nature (Kurakin, 2011). Somehow, the code of nature is such, that distinct systems (even social ones) self-organise into similarly structured patterns. As an alternative explanation for universal patterns, we can take a more Kantian perspective and question the objective nature of the patterns we observe. As stated before, patterns are abstract structures. What we consider to be structured and unstructured may be shaped by our psychology and limited by our inability to grasp the complexity of the world. This is in line with notion from Gestalt Theory such as presented in Palmer (1999). Human psychology has a tendency to structure pieces of information into larger information structures in certain ways. The notion of universal patterns that I put forward here can be interpreted ontologically neutral. We are simply dealing with the observation that universal patterns are observed by scientists and thereby partially determine which models we consider to be useful.

4 Yule Process: A Case Study

Finally, to illustrate the account I have described above, I would like to discuss the Yule Process and the universal pattern that can be derived from it; the Yule distribution. I have chosen this example of model transfer, because there exists an explicit account of how this model has been constructed in Yule (1925) for its original context, as well as how the model structure was later used as a basis for the construction of models in other domains (Simon, 1955). More recently, the Yule Process has formed basis for many models that concern preferential attachment (Abbasi, Hossain, & Leydesdorff, 2012), which is a central notion in network theory (Newman, 2001).

4.1 Yule Process: Evolutionary Origins

George Undy Yule (1871-1951) is known as a pioneer in the field of statistics. The model that is the subject of this case study is called the Yule Process. The distribution that can be derived from this process has been labelled the Yule Distribution, which is perhaps his most well-known scientific contribution (Edwards, 2001). A short history of the development of the model can be found in Bacaër (2011), on which the analysis below is partially based.

Yule developed his model in response to observations made by botanist J.C. Willis (1868-1958) in evolutionary biology. The issue concerns the distribution observed in taxonomy. Taxonomy is a biological classification scheme with a hierarchical structure in which organisms are grouped together based on common characteristics. The system is hierarchical in the sense that classifications with a so-called higher taxonomic rank are more general, and, thus, embed a classification of more specific lower taxonomic ranks. The observations made by Willis regards two such ranks, specie, and the more general rank of genus. A given genus thus contains multiple species, which have some features in common at the genus level but differ at the species level. The suborder of -Snakes-, for example, contains many more specific genera such as -Boa- which, in turn, contains the specie of -Boa Constrictor-. For several different organisms, animals and plants, Willis collected data on the number of genera that contain a given number of species. In this context, we can say that the size of a genus is determined by the number of species it contains. By tabulating this data, an interesting distribution emerged; there are many genera that contain one specie (size one), there were some larger genera, and some genera that were very large and contained more than a 100 species (size 100). What was also striking, is that this pattern appeared to emerge both in animals and plants. Yule, who was trained as a statistician under Karl Pearson, suggested to plot the data on a log-log scale. This revealed that the logarithm of the fraction of genera containing k species, $log(p_k)$, decreased approximately linearly with $log(k)$. This implies that there exists $\alpha > 0$ and $\beta > 0$ such that the probability density function of genera size can be written as:

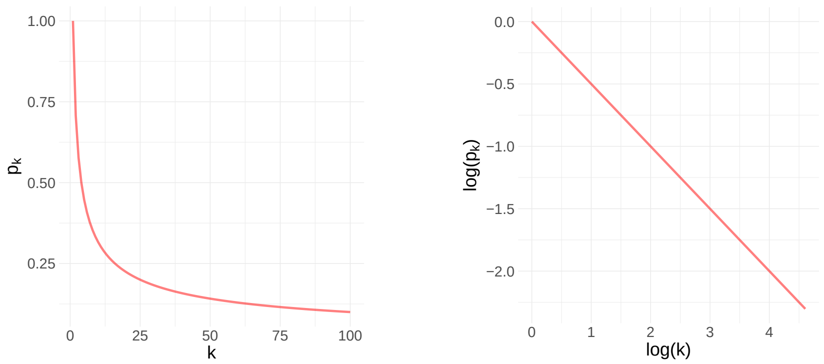
$$p_k \propto \alpha k^{-\beta} \tag{1}$$

Which can be rewritten as:

$$\log p_k \propto \log(\alpha) - \beta \log k \tag{2}$$

In Figure 2, I have plotted both equations for arbitrary parameters. In addition, J.C. Willis made observations regarding the age of a genus and its size. Stating that larger genera were on average older, evolutionary speaking.

Yule was interested in providing a mathematical model, based on evolutionary



(a) $p_k = 0.5k^{-1}$ (b) $\log p_k = \log(0.5) - \log k$

Fig. 2: Power Law for $\alpha = 0.5$ and $\beta = 1$

theory, that was able to reproduce (1) and, in addition, to explain the observation made by Willis that the larger genera were also older. In Yule (1925) he provided this model. Yule stated the purpose of his model as follows:

The Further question arises, what is the frequency distribution, as the statistician terms it, of the sizes of these N genera which all started as monotonic genera from primordial species at zero time, after any given time has elapsed?(Yule, 1925)

This purpose encapsulated the desire to generate the distribution of genera size as well as linking genera size to evolutionary age. From the outset, there were thus some clear validation criteria, that are in line with the ones I have discussed. There was a theoretical criterion, in that the model assumptions must roughly agree with evolutionary theory, and, there was a more explicitly phenomenological criterion: the model must able to reproduce a distribution that is linear on a log-log scale.

Let us now take a look at how Yule managed to construct a model that reproduces a frequency distribution that is in agreement with these "known facts". The two fundamental entities in this model are species and the genera they belong to. We consider how these two entities grow over time. The total number of genera is labelled as n. Each genus has a size k that is determined by the the number of species belonging to each genus at a point in time. In each time step, m species in total are added to the existing genera. After these m species have been added a new genus is added to the existing genera. This new genus starts out with $k = 1$. After this, the total number of species has thus increased by $m + 1$ (m plus the specie that is associated with the new genus). $m + 1$ new species appear for each new genus that is added, implying that the average number of species per genus is $m + 1$. With each time step

n is increased by 1. This implies that the number of time steps can be represented by the total number of genera n. $p_{k,n}$ is the fraction of genera with k species when the total number of genera is n. The total number of genera with k at n is $np_{k,n}$. Crucial now, is the probability of a species being added to an existing genus. This probability is taken to be proportional to the size of the genus, such that, if we have a genus with k_i species the probability of a specie to be added to this genus is given by the number of species belonging to genus i over the total number of species.:

$$\frac{k_i}{n(m+1)}. \tag{3}$$

We now have all the ingredients of the model. In short, the model consists of two main elements; constant genera growth and proportional specie growth. The question to ask is where do these ingredients come from? Part of it is a general knowledge of evolutionary theory. In the introduction to his paper, Yule discusses two opposing views regarding how evolution occurs that were relevant during his time. First is what Yule labels as the "Darwinian view", which assumes that differences in species and genera arise through cumulative small mutations (continuous variation) and that species necessarily die out. The "mutational view" assumes that large mutations may occur "at once *per saltum*", as Yule phrases it, which means with large jumps (discontinuous variation). It may seem that the type of mutation described in the model as well as the assumption that species do not die out, is more in line with Mutationalism. Yule is well known for his opposition to Mutationalism, which is most prominently featured in Yule (1902). In turn, to ensure that his assumptions do not disagree with the Darwinian view, Yule provides us with an explanation of how the model's assumptions should be interpreted. First, mutations in his model are limited to "viable mutations", such that the model does not formally contract the dying out of species. Second, Yule points out that given a long enough time horizon, small continuous mutations accumulate to changes that may appear as discontinuous. The time horizon in the model should thus be interpreted as long enough for such small mutations to accumulate to something that would be classified as a new specie or a new genus. There was thus a clear effort to position the model within the context of existing evolutionary theory. Such considerations provide us with an example of how the ability to meet theoretical criteria are built into the structure of the model.

The model proposed by Yule, however, was certainly not a one-to-one mapping of evolutionary theory. Interestingly, behind proportional growth is the assumption that the probability of creating a new specie is the same for each individual species regardless of genus and time. This implies that larger genera will grow at a higher rate in absolute terms. Regarding this assumption, Yule states:

The assumption that the chances of specific (or generic) mutation are identical for all forms within the group considered are constant for all time are unlikely to be in accordance with the facts, but have to be made to simplify the work. (Yule, 1925)

Why did Yule make this non-factual assumption? Here we enter analytical tractability/mathematical criteria: Introducing heterogeneity in the rates at which hundreds of species and genera evolve would undoubtedly complicate the model's computational structure, and might hamper the degree to which the model would enhance understanding. In addition, it could be that such a model can only be implemented through computer simulation, which was not a tool available to Yule. To convince the reader about the correctness of this assumption, Yule points not to evolutionary theory but to empirical facts that the model must be able to reproduce, the phenomenological criteria:

In so far as the deductions do not agree with known facts the assumptions are probably incorrect or incomplete. In so far as we find agreement, or the more nearly we find the agreement, the assumptions are probably correct.(Yule, 1925)

The model proposed by Yule indeed is able to reproduce the frequency distribution of genera:

So for as the graphic test goes, accordingly, the theory gives very well indeed precisely the form of the distribution required.(Yule, 1925)

From the outset, before any formal derivation, we can see that the constant addition of small genera, coupled with a proportional growth of species would generate a distribution with some very large genera and many smaller ones. To put it mathematically, a skewed distribution. Starting with only genera with $k = 1$, some genera, by chance, will grow slightly larger than others. These larger genera will then have a higher probability of growing even larger (following equation (1)) and so on.

The description of the construction of the Yule Process shows how the model structure is shaped by a balancing act between three validation criteria: The model had to some extent be in line with notions from evolutionary theory, the model had to be solvable analytically, and the model needed to reproduce the observed statistical distribution. It it these criteria that served as the standards for model usefulness to Yule. This shows that the Yule Process is a model that was constructed for a specific domain and the structure is shaped by the validation criteria within this domain.

4.2 The Yule Process as a Model for Firm Growth

How was the structure of the Yule Process, a model constructed and used as a basis for the construction of models in other domains? In the analysis we have established that overlap in validation criteria between domains is necessary for models to be useful in multiple domains. Let us look, therefore, at which considerations were most important in the selection of the Yule Process as a basis for constructing models in a new domain.

The Yule Process has been used to model processes of many different subjects (Simon, 1955). As an example, we will look at how the Yule Process was first applied to model the distribution of firm size in Simon and Bonini (1958). Let me first provide a little background of the scientific discussions regarding models of firm size at the time of Simon and Bonini (1958). At that time, it had long been observed that the distribution of firm size is heavily skewed (Gibrat, 1931), implying a distribution in which there are some very large firms and many smaller firms. The non-normality of this distribution was seen as evidence of the non-trivial nature of the growth process. The observation brought with it, a dissatisfaction of standard economic theory because it was unable to make predictions regarding the distribution of firm size (Simon & Bonini, 1958). Born from this dissatisfaction, the goal in Simon and Bonini (1958) was to provide a model that was able to generate the observed distribution of firm size. From the start, the model construction was thus aimed at a phenomenological criterion.

Simon and Bonini (1958) starts with the assertion that in order to generate the distribution of the type observed in firm size, the law of proportional effect is an essential ingredient for the model. The law of proportional effect was first introduced by Gibrat (1931) and entails that growth is proportional to size. It is the same structure labelled by Yule as proportional growth. In the case of firms, this would mean that the same percentage of growth rates applies to firms of different sizes. This implies that larger firms grow faster in absolute terms. Concretely, this means that the expected percentage return on investments is not a function of firm size. Computationally, this is in line with growth in the original Yule process, in which larger genera will grow at higher absolute rates as well. This, however, was not enough to narrow down the appropriate model to one. Simon and Bonini (1958) states that there may be multiple distinct growth processes (model structures) that will generate the type of distribution skewness observed empirically as long as proportional growth is incorporated:

If we incorporate the law of proportionate effect in the transition matrix of a stochastic process, then, for any reasonable range of assumptions, the resulting steady-state distribution of the process will be a highly skewed distribution, much like the skewed distribution of that have been so often observed for economic variates. In fact, by introducing some simple variations into the assumptions of the stochastic model - but retaining the law of proportionate effect as a central feature of it - we can generate the log-normal distribution, the Pareto distribution, the Yule distribution, Fisher's log distribution and *others - all bearing a family resemblance through their skewness.*(Simon & Bonini, 1958)

Proportional growth was thus deemed as essential for generating the type of distribution that was observed for the size of firms. This still left open, however, a range of skewed distributions and processes that generate them. In order to narrow down the growth process further, Simon and Bonini (1958) looked more closely to the characteristics of the observed distribution of firm size.

The log-normal function has most often been fitted to the data and generally fits quite well. It has usually been noticed, however, that the observed frequencies exceed the theoretical in the upper tail and that the Pareto distribution fits better than the log-normal in that region. The observation suggests that the stochastic mechanisms proposed in the previous section are the appropriate ones and that the data should be fitted with the Yule Distribution.(Simon & Bonini, 1958)

The observed pattern is thus one of a particular shape: it is log-normal except for the upper tail which is Pareto distributed. These two characteristics are consistent with the pattern of the Yule distribution. In order to reproduce this pattern, Simon and Bonini (1958) incorporates the second essential ingredient of the Yule Process; constant entry of new small firms. In this way Simon and Bonini (1958) arrives at a model which has the same structure as the the original model and is able to meet the validation criteria within the new domain.

4.3 Overlapping Validation Criteria

Where can we find overlap in the validation criteria between the original and new domain? First, if we look at theoretical criteria, we do not see strong indications of overlap. The evolutionary theory that served as a criterion in the original construction of the Yule Process did not play an explicit role when the model structure was applied to firms. In Simon and Bonini (1958) we see that theoretical criteria did not seem to play a big role altogether. Rather, Simon and Bonini (1958) is partially born out of a dissatisfaction with the inability of microeconomic theory to explain empirical patterns. Second, for both models there was an, at least implicit, mathematical criterion of analytical tractability. The Yule Process was a good candidate because the model structure was shown by Yule (1925) to fulfil this criterion. In line with Knuuttila and Loettgers (2020), this criterion is fulfilled by countless model structures and is not enough to narrow things down to a particular model structure. By itself, it is not a complete explanation as to why the Yule Process was transferred to the new domain. Third, is the overlap between the pattern observed in the distribution of genera size and the pattern observed in the distribution of firm size. It was this pattern, a certain shape, that enabled the model structure of the Yule Process to be considered as useful in both domains.

5 Conclusion

What explains inter-domain model transfer in science? I have put forward an account of model transfer that starts from the construction process of models in practice. In practice, models are constructed such that they meet relevant validation criteria. These criteria can be theoretical, mathematical or phenomenological in nature. The structure of the models is shaped by these criteria. In this sense, a model structure can thus be seen as an artefact that meets certain criteria. If such criteria are domain specific, the model structure will only transfer within the original domain of construction. If, however, the validation criteria also apply to other domains to a large enough extent, the model structure may be considered a useful tool in these domains as well. Inter-domain overlap in theoretical criteria applies in cases where the core of the theory in question is sufficiently abstract, such as complexity science. Mathematical criteria play an important role in shaping many model structures and these criteria will often overlap between domains, analytical tractability, for example. I agree with Knuuttila and Loettgers (2020), however, that such criteria are in some sense so general that they to not constitute a complete explanation. They do not explain the fact that some particular model structures are transferred and others are not. Phenomenological criteria, in the form of an ability to reproduce certain patterns may overlap across domains if the pattern in universal. Universal patterns are abstract structures that can be fitted to facts about phenomena in multiple domains by coupling it with domain-specific empirical content. Why we observe such patterns in an ontological question which may tell us something about how nature self-organises into typical structures, or may tell us something about our way of dealing with the limitations of grasping nature's complexity.

The case of the Yule process provides us with evidence that universal patterns are what enables model transfer in some instances. The case shows how a the Yule distribution shaped the original Yule Process model to a large degree. Stripped from its ontological content, the Yule Process is a device that generates a specific pattern in an analytically tractable way. The reason why Simon and Bonini (1958) uses the same model structure to constructed a model of firm growth is clear; the model structure was able to reproduce a specific pattern. It was this phenomenological validation criterion that enabled the model transfer. Importantly, the pattern is the starting point for Simon and Bonini (1958), and not the way in which the mechanisms of the model, proportional growth and constant addition of new entities, could be made to apply to firms instead of genera.

The Yule Process case study, presents us with an instance in which overlap in phenomenological criteria was the primary reason that the particular model structure of the Yule Process was transferred between domains. It is important to state, however, that in other cases (for example Knuuttila and Loettgers (2020)), the primary reason for model transfer may overlap in theoretical and/or mathematical criteria.

The added value of the account presented in this paper is threefold. First, instead of starting from a particular epistemological view regarding what makes models useful, it starts from looking at how models are constructed in practice. In practice, it is validation process that determines when a model is considered to be useful. The account is, therefore, neutral in the sense that is open to a multitude of epistemological viewpoints. Whether we consider models to be close representations of the reality or more akin to measurement instruments, for example, ultimately depends on what it means that a model fulfils certain validation criteria. Second, by introducing the notion of overlap in phenomenological criteria as an enabling source of model transfer in addition to analytical tractability and theoretical concepts, the account in this paper, extends the account of the model template (Knuuttila & Loettgers, 2016) to apply to a wider variety of model transfer cases. Third, it provides a concept that answers to some degree why overlap in phenomenological criteria may occur or even be prevalent, namely universal patterns.

References

Abbasi, A., Hossain, L., Leydesdorff, L. (2012). Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics*, *6*(3), 403–412.

Bacaër, N. (2011). *A short history of mathematical population dynamics*. Springer Science & Business Media.

Barlas, Y. (1996). Formal aspects of model validity and validation in system dynamics. *System Dynamics Review: The Journal of the System Dynamics Society*, *12*(3), 183–210.

Batterman, R.W. (2000). Multiple realizability and universality. *The British Journal for the Philosophy of Science*, *51*(1), 115–145.

Batterman, R.W., & Rice, C.C. (2014). Minimal model explanations. *Philosophy of Science*, *81*(3), 349–376.

Boumans, M.J. (1999). Built-in justification. M.S. Morgan & M. Morrison (Eds.), *Models as mediators: Perspectives on natural and social science* (p. 66-96). Cambridge University Press.

Boumans, M.J. (2006). The difference between answering a 'why' question and answering a 'how much' question. *Simulation* (pp. 107–124). Springer.

Boumans, M.J. (2009). Understanding in economics: Gray-box models. H.W. De Regt, S. Leonelli, & K. Eigner (Eds.), *Scientific understanding: Philosophical perspectives* (pp. 210–229). University of Pittsburgh Press.

Corominas-Murtra, B., & Solé, R.V. (2010). Universality of zipf's law. *Physical Review E*, *82*(1), 011102.

Donhauser, J. (2020). Informative ecological models without ecological forces. *Synthese*, *197*(6), 2721–2743.

Dosi, G., & Nelson, R.R. (1994). An introduction to evolutionary theories in economics. *Journal of evolutionary economics*, *4*(3), 153–172.

Edwards, A. (2001). George Udny Yule. *Statisticians of the centuries* (pp. 292–294). Springer.

Gandolfo, G. (2008). Giuseppe palomba and the lotka-volterra equations. *Rendiconti Lincei*, *19*(4), 347–357.

Gatti, D.D., Fagiolo, G., Gallegati, M., Richiardi, M., Russo, A. (2018). *Agent-based models in economics: a toolkit.* Cambridge University Press.

Gibrat, R. (1931). *Les inégalites économiques.* Sirey.

Hesse, M. (1966). *Models and analogies in science.* University of Notre Dame Press.

Humphreys, P. (2004). *Extending ourselves: Computational science, empiricism, and scientific method.* Oxford University Press.

Humphreys, P. (2019). Knowledge transfer across scientific disciplines. *Studies in History and Philosophy of Science Part A*, *77*, 112–119.

Knuuttila, T., & Loettgers, A. (2016). Model templates within and between disciplines: from magnets to gases–and socio-economic systems. *European journal for philosophy of science*, *6*(3), 377–400.

Knuuttila, T., & Loettgers, A. (2020). Magnetized memories: Analogies and templates in model transfer. *Philosophical perspectives on the engineering approach in biology* (pp. 123–140). Routledge.

Kurakin, A. (2011). The self-organizing fractal theory as a universal discovery method: the phenomenon of life. *Theoretical Biology and Medical Modelling*, *8*(1), 1–66.

Lloyd, E.A. (2015). Model robustness as a confirmatory virtue: The case of climate science. *Studies in History and Philosophy of Science Part A*, *49*, 58–68.

Lyon, A. (2014). Why are normal distributions normal? *The British Journal for the Philosophy of Science*, *65*(3), 621–649.

Mandelbrot, B. (1982). *The fractal geometry of nature* (Vol. 1). WH freeman New York.

Mandelbrot, B., & Hudson, R.L. (2007). *The misbehavior of markets: A fractal view of financial turbulence.* Basic books.

Newman, M.E. (2001). Clustering and preferential attachment in growing networks. *Physical review E*, *64*(2), 025102.

Palmer, S.E. (1999). *Vision science: Photons to phenomenology.* MIT press.

Parunak, H.V.D., Brueckner, S., Savit, R. (2004). Universality in multi-agent systems. *Autonomous agents and multiagent systems, international joint conference on* (Vol. 3, pp. 930–937).

Senge, P.M., & Forrester, J.W. (1980). Tests for building confidence in system dynamics models. *System dynamics, TIMS studies in management sciences*, *14*, 209–228.

Simon, H.A. (1955). On a class of skew distribution functions. *Biometrika*, *42*(3/4), 425–440.

Simon, H.A., & Bonini, C.P. (1958). The size distribution of business firms. *The American economic review*, 607–617.

Van Fraassen, B.C., et al. (1980). *The scientific image.* Oxford University Press.

Yule, G.U. (1902). Mendel's laws and their probable relations to intra-racial heredity (continued). *New Phytologist*, *1*(10), 222–238.

Yule, G.U. (1925). A mathematical theory of evolution, based on the conclusions of Dr. JC Willis. *Philosophical transactions of the Royal Society of London. Series B*, *213*(402), 21–87.

Philosophy and Science: An Ontological Approach

Fabio Tononi

General Philosophy of Science

Philosophy and Science: An Ontological Approach

Short Abstract

Since the emergence of ancient Greek philosophy, the relationship between philosophy and science has varied considerably. For example, the idea of science in Scholasticism differs from that which took shape at the time of Galilei or in the period of the atomic physics of Bohr and Heisenberg. Today, different scholars suggest that philosophy has been replaced by science, whereas others do not see a real difference between the two. However, I argue that a distinction between philosophy and science, as well as an ontological definition, would be fruitful for both disciplines and would clarify a series of crucial epistemological issues.

Extended Abstract

Since the emergence of ancient Greek philosophy, the relationship between philosophy and science has varied considerably. For example, the idea of science (in relation to philosophy) in Scholasticism differs from that which took shape at the time of Galilei or in the period of the atomic physics of Bohr and Heisenberg. Today, different scholars suggest that philosophy has been replaced by science, whereas others do not see a real difference between the two. However, I argue that a distinction between philosophy (which is grounded on reason) and science (which is based on experiments and empirical data), as well as an ontological definition, would be fruitful for both disciplines and would clarify a series of crucial epistemological issues.

In recent times, a series of important scientific discoveries – particularly in the fields of quantum physics, quantum cosmology, and cognitive science – has encouraged scientists to address questions that were usually the exclusive domain of theology and philosophy. For example, What is the nature of reality? Did the universe need a creator? Do we have free will? Do we have a self? and so on.

In *The Grand Design* (2010), Hawking and Mlodinow boldly claim that 'philosophy is dead' and therefore it can no longer answer these questions. As they argue, 'Philosophy has not kept up with modern developments in science, particularly physics'. They continue, 'Scientists have become the bearers of the torch of discovery in our quest for knowledge'. Therefore, they hold that 'only recent discoveries and theoretical advances' are able to suggest philosophical answers.

In his *Philosophy: The Latest Answers to the Oldest Questions* (2005), Fearn is equally radical. He points to the gradual transformation of philosophical problems into scientific ones. In this sense, Fearn argues, 'The difference between philosophy and science is often a matter of timing rather than a division of subject matter. Sometimes philosophy terminates in science'.

To shed light on this issue, this study focuses on Heidegger's definitions of philosophy and science while also considering the following questions: (*i*) Why is it relevant to address this issue today? (*ii*) Does the current multidisciplinary approach – such as the dialogue between philosophy and science in experimental metaphysics – challenge the idea of redefining the notion and task(s) of philosophy? And (*iii*) what is the task of philosophy in an age in which science seems to triumph?

Heidegger argued that philosophy is *thinking* (and untimely) and science is *knowledge* (and contingent). Heidegger's definition of philosophy largely derives from his readings of Aristotle and Nietzsche. Other, more recent thinkers agree with Heidegger's definition, adding that science has no memory (Lacan) and that science is (also) capitalism (Žižek). Clearly, all of this has important consequences for our view of the world and thus is essential to assess.

To conclude, it has become frequent to regard philosophy as a discipline that terminates in science. Therefore, the question is: Does philosophy still make sense in an age when science seems to answer all questions empirically? As I argue in my paper, at least a couple of areas cannot be addressed empirically, but by reason alone. These are:

> (1) The ontological questions that investigate the essence of reality, or questions such as 'what is philosophy?' and 'What is science?'; and

(2) The ethical and moral questions that investigate what is good and evil in the community.

As I propose, we will never know the answers to questions of these kinds from a mere empirical experiment or mathematical calculation.

Keywords

Heidegger, ontology, philosophy, science

Measures for Fighting Linguistic Injustice: Epistemic Equity and Mitigation

Aleksandra Vučković and Vlasta Sikimić

A) General Philosophy of Science

Measures for Fighting Linguistic Injustice: Epistemic Equity and Mitigation

Keywords: linguistic injustice, science, equity measures, mitigating agents

In recent years, there has been a lot of discussion regarding *lingua franca* in scientific research. While having English as a shared language within the scientific community contributes to more efficient communication and peer reviews, it also puts non-native speakers in a disadvantaged position. They have to invest a lot of time and sometimes money into perfecting English, while native speakers have the privilege of not having to put in extra effort and, therefore, have more time for their research. Moreover, we argue that some concepts are inherently untranslatable and, thus, the pluralism of languages can prevent the loss of unique concepts. We use Quine's famous thesis on the indeterminacy of translation to demonstrate the way the knowledge of marginalized scientists gets lost. We analyze the consequences of linguistic injustice in science and propose two measures for overcoming it: practicing epistemic equity and introducing mitigating agents.

Following the initial dismissal of Nobel prize winner Harald zur Hausen's discovery that cervical cancer is caused by the HPV virus (Cornwall 2013), we explore several types of epistemic injustice. We argue that zur Hausen's team suffered linguistic testimonial injustice due to the language barrier which resulted in their findings being disregarded. Moreover, they also experienced hermeneutical injustice since one of the reasons for the initial disbelief was that their findings were unexpected.

The correlation between hermeneutical and testimonial injustice has been previously explored in philosophical research, as well as potential solutions. For instance, Anderson (2012) proposes that integration and equality should be considered central epistemic virtues of the scientific community. We would like to strengthen her thesis and advocate for the scientific policy based on the principle of equity. While this principle has been widely represented in the context of education, it has yet to be implemented in the context of scientific research.

It should be recognized that most scientific journals are published in English and, therefore, are much more accessible to native speakers. To achieve equality, certain steps need to be taken to compensate for the disadvantages of non-native speakers. These may include free proofreading in English, accessible translation services, acceptance of the papers that are not written in *lingua franca*, etc. The principle of equity should also be extended to the other types of epistemic injustice since they are intertwined. Finally, linguistic diversity should be recognized as beneficial both to individual researchers and to the scientific community as a whole. Through the inclusion of the concepts that are unique to languages other than *lingua franca*, the whole corpus of scientific knowledge is enriched.

Moreover, mitigating agents are helpful for overcoming linguistic injustice in science. The main role of mitigating agents is the promotion and translation of scientific notions from various cultural and linguistic backgrounds. The task of mitigation can be taken by any scholar that is proficient in *lingua franca* and familiar with the topics of research. This process should create a bridge between marginalized researchers and the scientific community and needs to be done on several levels. From the global perspective, the research material needs to be available in as many languages as possible and the *lingua franca* spoken at the conferences should be adapted and simplified so that non-native speakers can understand it as well. On the individual level, scientists should practice epistemic openness to unusual concepts and the imperfect use of English.

References

Allison, P. D. (1980). "Inequality and scientific productivity". *Social Studies of Science* 10(2): 163-179.

Anderson, E. (2012). "Epistemic Justice as a Virtue of Social Institutions", *Social Epistemology*, Volume 26, 2012 - Issue 2: Epistemic Injustice: 163-173.

Beach, M.C., Saha, S., Park J., Taylor, J., Drew, P., Plank E., Cooper, L.A., Chee, B. (2021). "Testimonial Injustice: Linguistic Bias in the Medical Records of Black Patients and Women". *J Gen Intern Med*. 2021 Jun;36(6):1708-1714.

Buden, B., Nowotny, S., Simon, S., Bery A. & Cronin M. (2009). "Cultural translation: An introduction to the problem, and Responses". *Translation Studies*. Vol. 2, No. 2, 2009, 196-219

Butler J. (1996). "Universality in culture". In M.Nussbaum (Ed.) *For love of country? Debating the limits of patriotism*. Beacon Press: 45 – 52

Cornwall, C. (2013). *Catching cancer: the quest for its viral and bacterial causes*, Rowman & Littlefield.

Dotson, K. (2011). "Tracking epistemic violence, tracking practices of silencing". *Hypatia*, 26(2): 236-257.

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.

Grabe, W. (1988). "English, information management, and technology transfer: A rationale for English as an international language". *World Englishes*, 7: 63-72.

Huang, Junming, Alexander J. Gates, Roberta Sinatra, and Albert-László Barabási. 2020. "Historical Comparison of Gender Inequality in Scientific Careers across Countries and

Disciplines." *Proceedings of the National Academy of Sciences of the United States of America* 117 (9): 4609–16.

Hyland, K. (2016). "Academic publishing and the myth of linguistic injustice". *Journal of Second Language Writing* 31 (2016): 58-69

Jakobson, R. (1971). "On linguistic aspects of translation". *Word and language*. Vol. 2 of Selected writings. Mouton: 260-266.

Kitcher, P. (1990). "The division of cognitive labor". *The journal of philosophy*, 87(1), 5-22.

Koskinen, I., & Rolin, K. (2021). "Structural epistemic (in) justice in global contexts". In D.Ludwig, I.Koskinen, Z.Mncube, L.Poliseli & L.Reyes-Galidno (Eds.) *Global Epistemologies and Philosophies of Science*. Routledge: 115-125.

Leefman, J. (2021). "Social Exclusion, Epistemic Injustice, and Intellectual Self-Trust". *Social Epistemology*: 1-11.

Lefevere M. & Schliesser E. (2014). "Private Epistemic Virtue, Public Vices: Moral Responsibility in the Policy Sciences". In book: *Experts and Consensus in Social Science*: 275-295

Lillis, T. , Hewings, A. , Vladimirou, D. & Curry, M. J. (2010). "The geolinguistics of English as an academic lingua franca: citation practices across English‐medium national and English‐medium international journals." *International Journal of Applied Linguistics*, 20: 111-135.

Longino, H. (2001). *The fate of knowledge*. Princeton: Princeton University Press.

McGinnity, F., Nelson, J., Lunn, P. & Quinn, E. (2009). *Discrimination in Recruitment: Evidence from a Field Experiment*, Dublin: The Equality Authority.

Medgyes P. & Kaplan R.B. (1992). "Discourse in a foreign language: the example of Hungarian scholars". *International Journal of the Sociology of Language*. Vol. 1992 (Issue 98): 67-100.

Miller, B. (forthcoming). "Epistemic Equality". available at Epistemic Equality (Last Accessed: January 22nd 2022).

Mitova, V. (2020). Decolonising Knowledge Here and Now. Philosophical Papers, 49(2), 191-212.

Muresan L.-M. & Pérez-Llantada C. (2014). "English for research publication and dissemination in bi-/multiliterate environments: The case of Romanian academics", *Journal of English for Academic Purposes*. Vol. 13: 53-64.

Parkin, D. M. & Bray, F. (2006). "The burden of HPV-related cancers". *Vaccine*, 24, S11-S25.

Patten A. (2009). "Survey Article: The Justification of Minority Language Rights". *The Journal of Political Philosophy*: Volume 17, Number 1, 2009: 102–128

Pennycook, A. (2012). “Lingua Francas as Language Ideologies”. In *English as an International Language in Asia: Implications for Language Education*. eds. Kirkpatrick & Sussex: 137-156. Springer

Penuel W.R., Watkins D.A. (2019), “Assessment to Promote Equity and Epistemic Justice: A Use-Case of a Research-Practice Partnership in Science Education”, *AAPSS*: 201-216.

Postma, D. (2016). “Open Access and Epistemic Equality”. *Education as Change*. Vol 20: 1-10.

Pronskikh, V. (2018). “Linguistic Privilege and Justice: What Can We Learn from STEM?”. *Philosophical Papers*. Volume 47, 2018 - Issue 1: Linguistic Justice and Analytic Philosophy: 71-92.

Quine, W.V.O. (1960, 2013). *Word and Object*. The MIT Press.

Rawls J. (1971, 1999). *A Theory of Justice*. Belknap Press

Salager-Meyer, F. (2014), “Writing and publishing in peripheral scholarly journals: How to enhance the global influence of multilingual scholars?” *Journal of English for Academic Purposes 13*: 78-82.

Samarin W.J. (1968), “Lingua francas of the world”. In *Readings in the sociology of language*. ed. J. A. Fishman: 660-672. The Hague: Mouton and Co.

Schliesser E. (2018). “On Philosophical Translator-Advocates and Linguistic Injustice”, *Philosophical Papers*, Vol. 47, No. 1: 93-121.

Sikimić V., (2022). How to Improve Research Funding in Academia? Lessons From the COVID-19 Crisis. *Frontiers in Research Metrics and Analytics*. 7:777781.

Sikimić, V., Nikitović, T., Vasić, M., & Subotić V. (2021). “Do Political Attitudes Matter for Epistemic Decisions of Scientists?” *Review of Philosophy and Psychology*. 12: 775-801.

Skutnabb-Kangas, T. (1988). “Multilingualism and the education of minority children”. In T. Skutnabb-Kangas & J. Cummins (Eds.), *Minority education: From shame to struggle* (pp.9-44). Clevedon, UK: Multilingual Matters.

Smart, J.J.C. (1968). “Quine’s Philosophy of Science”. *Synthese*. Vol. 19, No. 1/2 (Dec., 1968): 3-13.

Straßer, C., Šešelja D., and J.W. Wieland. (2015). “Withstanding tensions: Scientific disagreement and epistemic tolerance”. In *Heuristic reasoning. Studies in Applied Philosophy, Epistemology and Rational Ethics*. ed. E. Ippoliti, vol. 16: 113-146. Cham: Springer.

Tardy, C. (2004). “The role of English in scientific communication: lingua franca or Tyrannosaurus rex?”, *Journal of English for Academic Purposes*, Vol. 3: 247-269.

Van Parijs, P. (2002). “Linguistic justice”. *Politics, Philosophy & Economics* 1.1: 59-74

Van Parijs, P. (2007). “Linguistic diversity as curse and as by-product” in *Respecting Linguistic Diversity in the European Union*. ed. Xabier Arzoz: 17-46 John Benjamins Publishing Company. Amsterdam.

Van Parijs, P. (2011). *Linguistic Justice for Europe and for the World*. Oxford University Press.

Visle, L. (2003). “From Integration to Inclusion: Focusing Global Trends and Changes in the Western European Societies”. *European Journal of Special Needs Education* 18 (1): 17–35.

Wang F. (2008). “Nationalism without Linguism? Reevaluating the Chinese orthography in the context of language revitalization”. 25-49.

Wickström, B. A., Templin, T., & Gazzola, M. (2018). An economics approach to language policy and linguistic justice. In *Language policy and linguistic justice* (pp. 3-64). Springer, Cham.

Yamada, S., Cappadocia, M. C., & Pepler, D. (2014). Workplace bullying in Canadian graduate psychology programs: Student perspectives of student–supervisor relationships. *Training and Education in Professional Psychology*, *8*(1), 58.

Zollman, K. J. (2007). “The communication structure of epistemic communities”. *Philosophy of science*. 74(5), 574-587.

Zollman, K. J. (2010). “The epistemic benefit of transient diversity”. *Erkenntnis*. 72(1), 17.

Short abstract:

While having English as a shared language within the scientific community contributes to more efficient communication and peer reviews, it also puts non-native speakers in a disadvantaged position. Moreover, we argue that some concepts are inherently untranslatable and, thus, the pluralism of languages can prevent the loss of unique concepts. We use Quine's thesis on the indeterminacy of translation to demonstrate the way the knowledge of marginalized scientists gets lost. We analyze the consequences of linguistic injustice in science and propose two measures for overcoming it: practicing epistemic equity and introducing mitigating agents.

Why Is the Extended Mind a Misleading Case? Towards a Mechanistic Account of DCog

Witold Wachowski

Section: Philosophy of Cognitive and Behavioral Sciences

Keywords: distributed cognition; heuristics; integration; mechanistic explanation; wide cognition

Title: Why Is the Extended Mind a Misleading Case? Towards a Mechanistic Account of DCog

Extended abstract [978 words]:

The main aim of my research is to elaborate the theoretical lynchpin of distributed cognition research in the form of cognitive mechanisms, in a way that integrates research on so called wide cognition (cases and types of cognitive processes that cannot be reduced to the activity of individual brain) with the rest of cognitive science. This is the answer to a persistent, complex problem in the field. This paper refers to one of my research steps, which is the distinction of two dimensions of the distributed cognition approach: a theoretical framework and a task model. In my project, I focus on the former.

At the time of its founding, cognitive science seemed to sufficiently appreciate the sub-disciplines co-constituting it. It took into account the socio-cultural aspects of cognitive processes, which is why anthropology itself was in quite good relations with cognitive sciences at the time. Allen Newell pointed to the social aspects of cognition among his 13 criteria for unified cognitive theory. Donald Norman sees cultural knowledge systems among 12 key issues for cognitive sciences. Howard Gardner listed affect, context, culture, and history as important cognitive components. Finally, however, the participation of anthropology was not sufficient despite favorable conditions. The role of society and culture has been eliminated from the heart of the issue of cognitive sciences, which was also due to the Newell's followers in creating a unified theory of cognition. Issues of the mainstream of cognitive sciences were strongly informed by methodological individualism (the approach according to which the study of the human individual is both necessary and sufficient to learn all the important aspects of cognitive processes), which has not changed at the stage of integration of psychology with neuroscience, regardless of the mechanistic dimension of this integration.

In the early 1990s, lively discussions on the embodied mind and situated cognition began and continue to this day (see 4E approach). It seemed that there was a new, logical stage of enriching and extending the scope of cognitive science. Basically, the new trends were coupled with the critique of traditional approaches in cognitive science, including a rather hasty reduction or even negation of the importance of computationalism and cognitive representationism. Contrary to the intentions of their representatives, the concepts of wide cognition, although already present in the mainstream of cognitive sciences, are not intended to merge or integrate, but rather widen the gap between them and what is generally recognized as specific to cognitive science: the assumption of the computational nature of cognitive processes, comparing natural intelligence with artificial one, combining formal and

empirical methods, or the use of computer simulations.

In cognitive science, the tendency to mechanistic integration can be observed with a wider dimension than the "narrow" integration initiated in the late 1980s and focused on neuroscience. Advocates of this "wide" integration (see Miłkowski et al., 2018) point out that individual concepts of wide cognition, contrary to the hasty describing them as "theories" or even "conceptual frameworks," yield only a fragmentary picture of wide cognition, without detailed predictions about to the phenomena described. Rather, they are research traditions that provide important heuristics for mechanism-based explanations, which enables slow evolution towards their integration with classical cognitive science focused on neuroscience. The wide perspectives on cognition seem to be fruitful when applied together in the practice of building mechanistic models.

Among the concepts of wide cognition, the distributed cognition approach (DCog) deserves special attention. It fully implements the basic heuristics of the ecological research tradition. According to this heuristics, socio-cultural conditioning has to be taken into account not at a later stage of analysis, but at its beginning. As Nancy Nersessian points out, the view dominant for a long time "has mistakenly attributed the properties of a complex, cognitive system, comprising both the individual and the environment, for the properties of an individual mind" (2009, p. 132). Moreover, DCog plays a unique role vis-à-vis other concepts of wide cognition. On the one hand, it proposes an understanding of cognitive processes and systems that most consistently breaks with methodological individualism and subject-focused approach. On the other hand, it demonstrates embedding in classical cognitive science by using the computational model of cognition and the concept of representation. Finally, according to Edwin Hutchins, the main representative of this approach, "the boundaries of the unit of analysis for DCog are not fixed in advance; they depend on the scale of the system under investigation, which can vary" (2014, p. 36).

In this context, it is possible to revise the critique of the approaches to wide cognition, showing to what extent the objections against DCog are unsuccessful, because they result from reducing it in practice to the extended mind approach. The cognitive process does not begin at any particular point in order to extend. DCog – in its most universalist interpretation – offers its own, yet underdeveloped, concept of integrating various cognitive research, regardless of the types and cases of cognitive activity, cancelling controversy between classical and non-classical approaches. This allows to see the relationship between DCog and the (neo)mechanistic explanation common in cognitive science.

Selected bibliography:

Anderson J.R. & Lebiere C. (2003). The Newell Test for a theory of cognition. Behav Brain Sci. 26(5).

Bender, A., Hutchins, E. & Medin, D. (2010). Anthropology in cognitive science. Top Cogn Sci . 2(3).

Boone, W. & Piccinini, G. (2016). The cognitive neuroscience revolution. Synthese 193(5).

Glennan, S. (2019). The New Mechanical Philosophy. OUP.

Hutchins, E. (1995). Cognition in the wild. MIT.

Hutchins, E. (2014). The cultural ecosystem of human cognition. Philos. Psychol. 27(1).

Miłkowski, M. et al. (2018). From Wide Cognition to Mechanisms: A Silent Revolution. Front. Psychol. 9, 2393.

Nersessian, N. J. (2009). Conceptual Change: Creativity, Cognition, and Culture. In Models of Discovery and Creativity, Meheus, J., Nickles, T. (Eds.). Springer.

Newell, A. (1990) Unified theories of cognition. HUP.

Osbeck, L. M. i Nersessian, N. J. (2014). Situating distributed cognition. Philos. Psychol. 27(1).

Short abstract [97 words]:

This paper refers to my research on distributed cognition (DCog) in the context of mechanistic integration in cognitive science. I point out the relationships and differences between DCog and the assumptions of the traditional cognitive science. I analyze the criticism against the wide cognition approaches, showing to what extent the objections against DCog are unsuccessful because they result from reducing it in practice to the extended mind approach. I also point to the relationship between DCog and the mechanistic explanation common in cognitive science, which allows to explain the role of ecological heuristics in research on cognition.

A Virtue Epistemology of Scientific Explanation and Understanding

Haomiao Yu

A Virtue Epistemology of Scientific Explanation and Understanding

Short Abstract: In this paper, I aim to develop a virtue epistemological account of scientific explanation and understanding. In so doing, I build a link between intellectual virtue and scientific explanation through understanding. The central epistemological question I will focus on is *how human beings understand the world through scientific explanation*. The answer I will give is that our understanding of the world is achieved by *the alignment of intellectual virtue and explanation structure*.

Extended Abstract:

There are various kinds of epistemology and different accounts of scientific explanation and understanding. The aim of this paper is to find a common ground on which philosophers' theory of knowledge agree with philosophers of science's theory of scientific explanation and understanding. This unification brings harmony to three lines of literature and settles their debates such as what the nature of understanding is.

Broadly speaking, there are currently two main approaches to the concept of intellectual virtue in the literature on virtue epistemology — virtue reliabilism and virtue responsibilism. They differ over the definition of virtue. Virtue reliabilists identify virtues as *reliable cognitive abilities or faculties* such as reason, perception, introspection, and memory etc. Virtue responsibilists, on the other hand, characterize intellectual virtues as *personality or character traits*, such as open-mindedness and fair-mindedness. In this paper I side with the (reliabilism-centered) reconciling strategy in treating reliabilist faculty virtues as *fundamental/constitutive* and responsibilist character virtues as *auxiliary.* I further argue that the reliabilist knowledge-constitutive virtues are fundamental in producing scientific knowledge, and they do so by mapping the virtues onto *the explanation structures* identified in the

literature on scientific explanation.

Explanation structures exhibit human reasoning abilities and skills, such as *inferential reasoning* in the structure of scientific deduction and induction, *causal reasoning* in the structure of causal explanation, *mathematical reasoning and skills* in the structure of mathematical explanation, etc. Hence, the reliabilist knowledge-constitutive virtues are *grounds on which scientific explanations construct*. They are the bones of scientific explanation. They thereby prove to be fundamental/constitutive in forming scientific knowledge.

Despite resistance from Hempel, discussions of *understanding* enter the field of scientific explanation. At the same time, virtue epistemologists also treat understanding as cognitive achievement that is as valuable as knowledge. There are three accounts of understanding in the literature of philosophy of science and virtue epistemology: Knowledge vs Ability vs Cognitive Achievement. I argue that understanding should be defined as a cognitive achievement, against the classic ability account and the newer knowledge account. This way, we will be able to find a common ground for both philosophy of science and virtue epistemology.

My central claim: The production of understanding is achieved by the alignment of intellectual virtue and explanation structure. The alignment is as follows:

Intellectual virtue		Explanation structure
Faculty or Ability	Deductive/Inductive reasoning	Deduction and induction
	Causal/mechanical reasoning	Causation/mechanism
	Counterfactual reasoning	Difference-making
	Generalization and categorization	Unification/consilience
	Approximation, abstraction and simplification	Idealization

Skill	Perception/Visual skills	Visualization
	Mathematical skills	Mathematical structures
	Model-building skills	Models
	Statistical skills	Statistical structures

How understanding is achieved? This can be illustrated by showing how the alignment will serve as the measure for both *degrees* and *kinds* of understanding. It's measured by different levels and kinds of alignment between intellectual virtue and explanation structure, regulated in terms of *types and tokens*.

To be specific, different types/kinds of understanding are represented by different kinds of alignment. For example, causal understanding (or understanding of causes) results from the alignment of causal reasoning and causal explanation, such as smoking causes cancer death. Within a single type/kind of understanding, namely a single kind of alignment, more alignment tokens represent deeper understanding, such as smoking causes lung cancer and lung cancer causes cancer death.

What happens when intellectual virtues don't align with explanation structures? Understanding isn't achieved. How does it happen? It could happen due to (1) *a lack of intellectual virtue,* (2) *a lack of explanation structure, or* (3) *a mismatch between intellectual virtue and explanation structure*.

For an illustration in the history of scientific practice, consider Galileo's experiments with pendulums. Here is a sketch of how knowledge is produced by virtues: It starts with Galileo's observation that the period the of the pendulum is dependent on its length, which is produced by intellectual virtues — *perception and inferential reasoning*. Then he offers support for it by means of the first law of astronomy", the process of which is carried through more intellectual virtues —

mathematical reasoning and the ability to generalize and formalize. This shows that Galileo's case falls in line with the tradition of virtue reliabilism that knowledge is produced by intellectual virtues (cognitive abilities or skills).

More importantly, Galileo's case is also a good illustration of my virtue account. With respect to explanation and understanding, we can draw two sets of conclusions regarding the virtue account:

(1) First, Galileo has the required mathematical reasoning abilities and skills, but Newton's laws are not available at the time; so, Galileo seeks theoretical support for his observation on pendulums from Kepler's laws. In this case, it *isn't* in nature a mismatch between intellectual virtue and explanation structure. Galileo's mathematical reasoning abilities and skills still *match onto* the mathematical structure of the pendulum. So, he can understand the isochrony of the pendulum, due to *the initial alignment* between intellectual virtue and explanation structure.

This is the first set of conclusions regarding explanation structures.

(2) However, Galileo's understanding of the isochrony of the pendulum is restricted to the mere correspondence of his mathematical reasoning ability with the mathematical structure of the pendulum, *provided by Kepler's law*. So, Galileo understands, *to some degree*, that the period is proportional to the square root of the length. Later, Newton formulates the laws of motion and universal gravitation, and he uses them to derive Kepler's laws. The pendulum law is thus formulated in terms of the gravitational constant. If Galileo were provided with Newton's work, his understanding would have been enhanced. That is to say, Galileo would have had a deeper understanding of the isochrony of the pendulum than his initial understanding based on Kepler's law.

Here, different degrees of the same type of understanding (mathematical understanding) are represented by more alignment tokens. Galileo's intellectual virtues aligning with Kepler's law results in a degree of mathematical understanding, namely one token of the mathematical type of understanding. But those virtues aligning with Newton's laws would result in a deeper understanding, another token understanding that is of the same kind — mathematical.

Moreover, a mismatch between intellectual virtue and explanation structure can also happen in Galileo's case. Since both Kepler's laws and Newton's laws are based on Euclidean geometry, Galileo would have been able to understand the pendulum through Newton's work via Euclidean model of reasoning. If Galileo were provided with the Schrödinger equation for the pendulum, he wouldn't have been able to understand the pendulum in a quantum-mechanical system, due to the lack of a proper training in quantum mechanics and the relevant mathematical skills. A mismatch thus happens, and understanding isn't achieved.

The above is the second set of conclusions regarding understanding.

Immunity in health and disease: a clash of frameworks

Martin Zach and Gregor Greslehner

b) Philosophy of Natural Science

Immunity in health and disease: a clash of frameworks

Keywords: immune system, defense, contextuality, regulation, trade-offs, strong immunity

Philosophy of immunology has grown into a small field within philosophy of science (see Pradeu [2019]; Swiatczak and Tauber [2020]). In fact, immunology has been investigated by philosophers in relation to a great many topics: the self/non-self theory of immunogenicity (Tauber [1994]; Pradeu [2012]), biological individuality and the related holobiont and ecological views (Pradeu [2016]; Schneider [2021]), the use of metaphors in biological, including immunological thinking (Martin [1994]; Tauber [1994]), and a variety of more specific notions such as immunological balance (Swiatczak [2013]).Here we contribute to the existing scholarship and consider a general framework (or account) of immunity. Our use of the terms 'framework' or 'account' indicates that we do not mean to propose a new *theory* of immunity (although there is a need for such a 'general theory of immunity', see, e.g., Eberl and Pradeu [2018]). Instead, we want to address the (still) widespread *mindset* from which one views the immune system, and we ultimately propose an alternative framework which better reflects recent advances.

The dominant characterization of the immune system found in immunological literature is that of a defense system which engages in 'strong' or 'weak' responses. We first scrutinize the mindset that the immune system is a defense system, and we conclude that such a framework fails to capture the general nature of the immune system because it omits immune functions and interactions that are unrelated to defense; similar arguments to that effect have been proposed by scholars before (Swiatczak [2013]; Tauber [2017]; Pradeu [2019]). Furthermore, even with defense in mind, there are other strategies than just the elimination of, e.g., pathogens, as showcased by disease tolerance (Medzhitov *et al.* [2012]). Second, thinking in terms of defense is tightly connected to the concepts of 'strong' and 'weak', which is why we subsequently focus on the immunological usage of these notions. We argue two things. First, interpreting quantitative measurements of immune responses in terms of 'strength' and 'weakness', despite being the most frequent immunological usage, does not add epistemic value to those measurements. Second, some immunological usage of the strong/weak framework calls for a broader consideration: we ask whether the strong/weak framework can shed light on the immune system and its activity generally. For that reason, we systematically consider a number of ways in which the notions of 'strength' and 'weakness' can be interpreted.

On the normative reading of these notions, one may associate strength with positive – and weakness with negative – connotations, respectively. However, such a picture turns out to be misleading, as 'strong' immunity or response is not necessarily desirable, and likewise 'weak' immunity is not always detrimental. Paradoxical connotation stems from the fact that an immune condition can oftentimes be viewed as both 'strong' and 'weak', and preferring one over the other is entirely arbitrary, thus uninformative. Given that the immune system is not monolithic, many particular functions may not be amenable to change, whereby the intuitive idea of making the immune system stronger or weaker breaks down. Moreover, many immunological phenomena and functions cannot be meaningfully captured by these notions. Finally, the strong/weak framework mischaracterizes the nature of the interactions between the immune system and other physiological systems, and what their respective contributions are. Therefore, we argue that the strong/weak framework fails at providing a general account of immunity.

Taken together, we argue that 'strong/weak immunity' and related notions are ill-defined and misleading. Although one can always develop a liberal enough interpretation of the terms such as 'defense' and 'strength' so that it fits *any* and *all* descriptions of observed phenomena, such a loose interpretation would render the notions devoid of meaning. As a result, we suggest that we need to move away from viewing the immune system narrowly as a defense system, one that could be accounted for in terms of 'strength' or 'weakness'.

After this critical assessment, we propose another framework which we harvest from the recent immunological literature – one that, once made explicit, provides insight into the general organizing principles of immunity. The framework we suggest as an alternative way to think about immunity provides core tools for framing, understanding, and studying the immune system. It emphasizes the crucial aspects of *co*ntextuality and *re*gulation of immunity, and the biological *t*rade-*o*ffs which the immune system exhibits. All immune-related phenomena require a contextual understanding; otherwise, one would fail to understand why a phenomenon may appear desirable in one context and detrimental in another. Regulation plays a paramount role in accounting for many ways in which the immune system operates or dysfunctions. Finally, one and the same component of the immune system that confers a particular benefit is also responsible for a poor outcome regarding another condition. Thus, the immune system exhibits numerous trade-offs.

Although the three concepts are well-known to immunologists, they usually appear as descriptors of discovered states of affairs. By explicitly analyzing these concepts, we propose that they should play a more prominent role in thinking about immunity. We also propose to unify these concepts into a single framework which we consider as a viable alternative to the problematic view of immunity as strong/weak defense. Finally, such a framework helps to achieve a better understanding of the organizing principles of immunity that allows addressing the role of the immune system in health and disease.

Eberl, G. and Pradeu, T. [2018]: 'Towards a General Theory of Immunity?', *Trends in Immunology*, **39**, pp. 261–3.

Martin, E. [1994]: 'Flexible Bodies: Tracking Immunity in American Culture from the Days of Polio to the Age of AIDS', Boston: Beacon Press.

Medzhitov, R., Schneider, D. S. and Soares, M. P. [2012]: 'Disease Tolerance as a Defense Strategy', *Science*, **335**, pp. 936–41.

Pradeu, T. [2012]: 'The Limits of the Self: Immunology and Biological Identity', Oxford: Oxford University Press.

——— [2016]: 'The Many Faces of Biological Individuality', *Biology and Philosophy*, **31**, pp. 761–73.

——— [2019]: 'Philosophy of Immunology', Cambridge: Cambridge University Press.

Schneider, T. [2021]: 'The Holobiont Self: Understanding Immunity in Context', *History and Philosophy of the Life Sciences*, **43**, pp. 1–23.

Swiatczak, B. [2013]: 'Immune Balance: The Development of the Idea and Its Applications', *Journal of the History of Biology*, **47**, pp. 411–42.

Swiatczak, B. and Tauber, A. I. [2020]: 'Philosophy of Immunology', in E. N. Zalta (*ed.*), *The Stanford Encyclopedia of Philosophy (Summer 2020 Edition)*,.

Tauber, A. I. [1994]: 'The Immune Self: Theory or Metaphor?', Cambridge: Cambridge University

Press.

——— [2017]: ‘Immunity: The Evolution of an Idea’, New York: Oxford University Press.

Can unconscious perception guide action?

Paweł Zięba

Can unconscious perception guide action?

1. According to unconscious perception hypothesis (UP) 'episodes of the same fundamental kind as episodes of conscious perception can occur unconsciously' (Block and Phillips 2017, 165). Phillips (Phillips 2018), the champion of scepticism about UP, argues that many putative instances of unconscious perception are cases in which unconscious perceptual representation of the stimulus is ill-suited to guide action. Consequently, there is no good reason to regard such representation as a personal rather than sub-personal state (i.e. to attribute it to the individual rather than to the individual's perceptual system), which means that it falls short of genuine, individual-level perception. This is the so-called 'problem of attribution'.

2. In this talk, I argue that the problem of attribution rests on unsound reasoning. The latter assumes that there is a sharp distinction between personal and sub-personal states/events and simultaneously violates that assumption by using personal-level criteria for perception and sub-personal-level criteria for action. Once perception and action are both identified in personal-level terms (as the assumption requires), the reason is lost to believe that the putative instances of unconscious perception cannot guide action.

3. According to Phillips, perceptual representation cannot guide action if it is unavailable to Central Coordinating Agency (a placeholder for 'whichever subsystems subserve an agent's genuine, individual-level action' (Phillips 2018, 497), from now on, CCA). Nevertheless, if one applies the distinction between personal and sub-personal states/events to perception, consistency requires that one applies it to all mental phenomena, including action. So instead of identifying action in relation to the sub-personal workings of CCA, Phillips should identify action in relation to the phenomenology of action and to the way we think and talk about action in everyday situations. But doing so renders the availability to CCA unnecessary for action. Many mundane activities we engage in in everyday situations are spontaneous, instinctive, and

1

occur without any rational deliberation. For example, jumping out of excitement occurs purely out of emotion (Hursthouse 1991), and tapping one's foot is not preceded by consciously intending or deciding to do so (Runyan 2014). None of possible outputs of CCA (e.g. conscious intention, conscious plan, conscious decision, conscious volition) seems necessary for those behaviours. And yet they are things we do; they don't just happen to us.

4. If action can occur without the involvement of CCA, unconscious perception doesn't have to be available to CCA in order to guide action. To reject the antecedent of this conditional, one has to either drop the distinction between personal and sub-personal states/events, or at least allow that the output of CCA can be unconscious. But doing so undermines the alleged contrast between conscious and unconscious perception that the personal/sub-personal distinction and the availability-to-CCA requirement were supposed to introduce. While it might be true that the activities mentioned in §3 involve some kind of unconscious volition or intention (and thereby require the output of CCA), this reply is unavailable to Phillips because it backfires on his scepticism about UP. For if the output of CCA can be unconscious, why think that the input of CCA has to be conscious? And if the input of CCA can be unconscious, why think that unconscious perception is unavailable to CCA?

5. My argument can be summarized as follows:

> 1. There is a sharp distinction between perception qua personal state/event and perception qua sub-personal state/event. [an assumption of the reasoning behind the problem of attribution]
>
> 2. The output of CCA (e.g. conscious decision, conscious intention, conscious plan, conscious volition) is necessary for action. [an assumption of the reasoning behind the problem of attribution]

3. There is a sharp distinction between personal and sub-personal states/events. [a general principle on which 1 is based]

4. There is a sharp distinction between action qua personal state/event and action qua sub-personal state/event. [from 3]

5. There are cases of action qua personal state/event that don't involve the output of CCA as defined in 2. [an assumption I shall motivate and defend from some objections]

6. The output of CCA as defined in 2 is *not* necessary for action. [from 5]

7. Premise 1 entails that Premise 6 is true [from 3, 4, 5], whereas Premise 2 entails that Premise 6 is false.

8. Premise 1 and Premise 2 cannot both be true. [from 7]

9. The problem of attribution rests on unsound reasoning. [from 8]

6. In conclusion, the reasoning behind the problem of attribution fails to establish that perception cannot guide action without the mediation of consciousness. This undermines one of the key components of Phillips' scepticism about UP.

References

Block, N., and I. Phillips. 2017. 'Debate on Unconscious Perception'. In *Current Controversies in Philosophy of Perception*, edited by B. Nanay, 165–92. New York, London: Routledge.

Hursthouse, R. 1991. 'Arational Actions'. *The Journal of Philosophy* 88 (2): 57–68.

Phillips, I. 2018. 'Unconscious Perception Reconsidered'. *Analytic Philosophy* 59 (4): 471–514.

Runyan, J.D. 2014. *Human Agency and Neural Causes*. Palgrave MacMillan.

Embryo-like structures, value-loaded metaphysics of science, and regulation of biomedical research

Tomasz Żuradzki

Section: Philosophy of natural science

Title: Embryo-like structures, value-loaded metaphysics of science, and regulation of biomedical research

Keywords: non-epistemic values, biomedicine, biological classifications, human embryos,

Philosophical discussions on non-epistemic values in science usually focus on the role of values in inspiring scientific questions, affecting scientific methodologies, and setting the level of evidence needed for drawing conclusions (Elliott, 2017). On the examples of human embryos and embryo-like structures (Ankeny, Munsie, & Leach, 2022), we analyze the role of non-epistemic values in classificatory practices in biomedicine. In particular, we focus on cases when regulatory mechanisms limit research, either in the forms of direct bans (e.g., some cases of human embryo research) or indirect incentives (e.g., bans on public funding or patenting some procedures regarding human embryos and embryo-like structures). Thus, we challenge all these philosophical theories of classification and kinds that do not accommodate the role for non-epistemic values (Khalidi 2013).

We will discuss the 14-day rule stating that in vitro research on human embryos and some embryo-like structures is permissible, but only until two weeks after fertilization or creation (Matthews & Moralí, 2020). One may interpret this rule as assuming that purely value-free biological facts about human embryos (e.g., individuation, i.e., the fact that embryos can no longer twin, or the first appearance of the primitive streak around this time, which is a precondition for the capacity to feel pain) ground the moral or legal status of organisms. Such a "metaphysics-first" approach tries to settle the metaphysical question of what a human embryo is – what is its essence or definition – and from there derive normative conclusions about, e.g., regulations on research (e.g. Lee 2004).

In this paper, we argue that this view is mistaken because, in particular, in the case of research-oriented biological classifications, there is no value-free (or interest-free) metaphysics of science (cf. Dasgupta, 2017). Our approach first takes into consideration the interplay between epistemic and non-epistemic values in real cases of biological classifications (Reydon & Ereshefsky, 2022), and then draws conclusions about metaphysics, i.e., the how human embryos or "synthetic human entities with embryo-like features" (Aach et al., 2017) may be classified to suit a given value framework.

For example, human parthenogenetic stem cells are excluded from the patenting prohibition of procedures based on hESC by the European Biopatent Directive, because such stem cells have been defined differently than human embryos or other types of stem cells: the parthenogenetic ones do not have 'the capacity' to develop into a (born) human being, i.e. totipotency (see European Court of

Justice 2014). However, the capacities of some of such embryo-like structures may be measured after realization of these capacities in controlled environments (Fagan, 2013). Since their capacity is context-dependent, there is no such thing as a value-neutral environment in which we can judge the embryo's or stem cells' 'genuine' intrinsic potential (Piotrowska, 2020). Thus, any forward-looking definition of embryos or embryo-like structures is grounded in specific normative evaluation on what counts as 'normal environment'.

We conclude with a few remarks on the role of philosophers of science and bioethicists in the realm of science policy and we argue that that major normative and regulatory issues in biomedical research would benefit from the tighter integration of these two disciplines (Lohse, Wasmer, & Reydon, 2020).

Representative references

Ankeny, R. A., Munsie, M. J., & Leach, J. (2022). Developing a reflexive, anticipatory, and deliberative approach to unanticipated discoveries: Ethical lessons from iBlastoids. *American Journal of Bioethics*, 22(1), 36-45.

Dasgupta, S. (2017). Essentialism and the Nonidentity Problem. *Philosophy and Phenomenological Research*, 46(3), 540-570.

Fagan, M. B. (2013). The stem cell uncertainty principle. *Philosophy of Science, 80*(5), 945-957.

Lohse, S., Wasmer, M. S., & Reydon, T. A. (2020). Integrating philosophy of science into research on ethical, legal and social issues in the life sciences. *Perspectives on Science, 28*(6), 700-736.

Reydon, T. A., & Ereshefsky, M. (2022). How to Incorporate Non-Epistemic Values into a Theory of Classification. *European Journal for Philosophy of Science*, 12(1), 1-28.

Symposia abstracts

Evidential Pluralism and its Application in the Social Sciences

Yafeng Shan, Jon Williamson and Alexandra Tromfimov

Title of the Corresponding Section

Philosophy of Social Sciences

Symposium's Title

Evidential Pluralism and its Application in the Social Sciences

List of Speakers

- Prof Jon Williamson (Department of Philosophy and Centre for Reasoning, University of Kent, UK) j.williamson@kent.ac.uk
- Dr Yafeng Shan (Department of Philosophy and Centre for Reasoning, University of Kent, UK) y.shan@kent.ac.uk
- Dr Alexandra Trofimov (Department of Philosophy and Centre for Reasoning, University of Kent, UK) a.trofimov-692@kent.ac.uk

Symposium Abstract

Evidential Pluralism is a normative thesis concerning the epistemology of causation. The basic idea of Evidential Pluralism is that in order to establish a causal claim that A causes B, one normally needs to establish the existence of an appropriate correlation between A and B and the existence of an appropriate mechanism complex linking A to B, so when assessing a causal claim one ought to consider both association studies and mechanistic studies, where available (Shan and Williamson 2021, 4).

Evidential Pluralism was originally introduced in the context of the health sciences (Russo and Williamson 2007) and has been fruitfully applied to the biomedical sciences (e.g. Gillies 2011; Clarke et al. 2014; Parkkinen et al. 2018; Williamson 2019; Canali 2019). However, the applicability of Evidential Pluralism in the social sciences has been controversial. For example, some (e.g. Weber 2009; Shan and Williamson 2021; Maziarz 2021) contend that Evidential Pluralism can be applied to the social sciences, while others (e.g. Reiss 2009; Claveau 2012; Beach 2021) are sceptical. This symposium examines the application of Evidential Pluralism to the social sciences.

Talk 1

Applying Evidential Pluralism in the Social Sciences

Evidential Pluralism maintains that in order to establish a causal claim one normally needs to establish the existence of an appropriate conditional correlation and the existence of an appropriate mechanism complex, so when assessing a causal claim one ought to consider both association studies and mechanistic studies. Hitherto, Evidential Pluralism has been applied to medicine, leading to the EBM+ programme, which recommends that evidence-based medicine should systematically evaluate mechanistic studies alongside clinical studies. This talk argues that Evidential Pluralism can also be fruitfully applied to the social sciences. In particular, Evidential Pluralism provides (i) a new approach to evidence-based policy; (ii) a new account of the evidential relationships in more theoretical research; and (iii) new philosophical motivation for mixed methods research. The application of Evidential Pluralism to the social sciences is also defended against two objections.

Talk 2

Evidential Pluralism and Political Science

According to Evidential Pluralism, in order to establish a causal claim that A causes B, one normally needs to establish the existence of an appropriate correlation between A and B and the existence of an appropriate mechanism complex linking A to B, so when assessing a causal claim one ought to consider both association studies and mechanistic studies, where available. In this talk, I shall argue for the application of Evidential Pluralism to political science. I shall argue that Evidential Pluralism can explain and validate successful causal analyses in political science, illustrated by Weinstein's study of wealth resources and violence in rebellions. Furthermore, I shall argue that Evidential Pluralism can make better sense of causal analysis in political science by providing a simple and unified epistemological account of causality and by shedding light on the roles of different methods in causal analysis.

Talk 3

EBL+: Applying Evidential Pluralism to Evidence Based Law

The emerging field of evidence based law (EBL) holds that law ought to be based on evidence rather than merely existing customs, ideals or morals. On the EBL approach, evidence is used to support a claim that a given law will adequately or optimally modify behaviour to achieve some desired end, such as reduced crime, increased safety or improvements in health.

Although disagreements concerning the nature or purpose of law present obstacles to the adoption of an evidence based approach, EBL is not without merit. Consider, for example, the introduction of a ban on using hand held devices while driving. Rather than justifying such a ban on common wisdom, we ought to have good evidence that such a ban would in fact reduce accidents and increase safety. Beyond the question of the appropriateness of an EBL approach in general, there is the question of what form an EBL approach ought to take. The efficacy of EBL depends, crucially, on what kind of evidence is required.

My aim in this talk is to motivate and defend an application of Evidential Pluralism (EP) to EBL. According to EP, establishing a causal claim requires evidence of both correlation *and* mechanism. Using the problem of online fake news as a case study, I argue that (i) an application of EP helps to overcome obstacles to establishing difference making in complex human behaviour and (ii) evidence of mechanism helps to identify, evaluate and justify effective legal interventions.

Jon Williamson

EMPLOYMENT

2005 - Philosophy, University of Kent. Professor of Reasoning, Inference and Scientific Method

2004 – 2005 Philosophy, London School of Economics. Tutorial Fellow.

1998 – 2004 Philosophy, King's College London. Research Fellow.

EDUCATION

1998 King's College London: PhD Philosophy.

1995 King's College London: MSc Philosophy of Science & Mathematics.

1994 Manchester University: BSc Hon's Mathematics.

SELECTED PUBLICATIONS

Books

Veli-Pekka Parkkinen, Christian Wallmann, Michael Wilde, Brendan Clarke, Phyllis Illari, Michael P. Kelly, Charles Norell, Federica Russo, Beth Shaw and Jon Williamson: *Evaluating evidence of mechanisms in medicine: Principles and procedures*, Springer, 2018.

Jon Williamson: *Lectures on inductive logic*, Oxford University Press, 2017.

Rolf Haenni, Jan-Willem Romeijn, Gregory Wheeler & Jon Williamson: *Probabilistic logics and probabilistic networks*, Synthese Library, Springer, 2011.

Jon Williamson: *In defence of objective Bayesianism*, Oxford University Press, 2010.

Jon Williamson: *Bayesian nets and causality: philosophical and computational foundations*, Oxford University Press, 2005.

Yafeng Shan

EMPLOYMENT

2019 – Research Associate, Department of Philosophy, University of Kent

2018 – 2019 IHPLS Research Fellow, Cohen Institute for the History and Philosophy of Science and Ideas, Tel Aviv University

2015 – 2018 Postdoctoral Researcher, Department of Philosophy, Durham University

EDUCATION

2016 PhD in Philosophy, University College London

2012 MLitt in Philosophy, University of St Andrews

2011 MSc in Philosophy of Science, London School of Economics

2009 BSc in Mathematics, University of Bristol

SELECTED PUBLICATIONS

Monograph

Shan, Yafeng. 2020. *Doing Integrated History and Philosophy of Science: A Case Study of the Origin of Genetics*. Boston Studies in the Philosophy and History of Science. Cham: Springer.

Journal Articles

Shan, Yafeng. 2022. "Philosophical Foundations of Mixed Methods Research." *Philosophy Compass* 17 (1): e12804.

Shan, Yafeng. 2021. "Beyond Mendelism and Biometry." *Studies in History and Philosophy of Science* 89: 155–63.

Shan, Yafeng, and Jon Williamson. 2021. "Applying Evidential Pluralism to the Social Sciences." *European Journal for Philosophy of Science* 11 (4).

Shan, Yafeng. 2020. "Kuhn's 'Wrong Turning' and Legacy Today." *Synthese* 197 (1): 381–406.

Shan, Yafeng. 2019. "A New Functional Approach to Scientific Progress." *Philosophy of Science* 86 (4): 739–58.

Alexandra Trofimov

EMPLOYMENT

2021 – Associate Lecturer, Kent Law School, University of Kent

2018 – 2021 Associate Lecturer/ Lecturer (Teaching and Scholarship), Philosophy, University of Kent

2013 – 2016 Graduate Teaching Assistant, University of Kent

EDUCATION

2017 PhD in Philosophy, University of Kent

2013 MA (Research) in Philosophy, University of Kent

2012 BA in Philosophy, University of Kent

AWARDS

- Keith Jones Prize for best performance at stage 3, 2012, University of Kent
- Above and Beyond Teaching Award from Kent Students' Union, 2019.

PUBLICATIONS

A. Trofimov, (forthcoming), 'Negligence is not Ignorance', *Jurisprudence*

A. Trofimov, (2016), 'Review of *The Continuum Companion to Ethics* by Christian Miller, *Journal of Moral Philosophy*, 13(1).

A. Trofimov (under review), 'Negligence Always Speaks Badly'

A. Trofimov (preparing for submission) 'What Ignorance Excuses'

Cognitive Philosophy of Science

Borut Trpin, Matteo De Benedetto, Nina Poth, Daniel Kostić and Mel Andrews

Cognitive Philosophy of Science
(symposium proposal)

Section: Philosophy of Cognitive and Behavioral Sciences

Talks and speakers (in order of presenting):

1. Matteo De Benedetto
   Ruhr-Universität Bochum
   matteo.debenedetto@ruhr-uni-bochum.de
2. Nina Poth
   Ruhr-Universität Bochum
   nina.poth@ruhr-uni-bochum.de
3. Daniel Kostić
   Radboud University
   daniel.kostic@gmail.com
4. Mel Andrews
   University of Cincinnati
   mel.andrews@tufts.edu
5. Borut Trpin
   MCMP/LMU Munich
   borut.trpin@lrz.uni-muenchen.de

Description

In its highly interdisciplinary approach, cognitive science studies the mind by employing concepts and methods from a variety of disciplines: psychology, artificial intelligence, neuroscience, linguistics, anthropology, and philosophy. Some of the unique insights provided by cognitive science have in turn influenced its foundational disciplines such as philosophy, e.g., the use of cognitive modeling in philosophy of science. However, a systematic philosophical study of these influences is still lacking in the literature.

We believe that the impact of cognitive science on philosophy of science is large enough to talk of a specific subfield, which we here tentatively call cognitive philosophy of science. We use this label to denote approaches in philosophy of science which are informed by cognitive science or use the methods, tools and research topics standard for cognitive science (or vice versa, how approaches in the philosophy of science inform cognitive science). Scientific reasoning, broadly conceived, is a standard umbrella term for the area which is addressed in cognitive philosophy of science.

Cognitive scientists also address various aspects of scientific reasoning (e.g., inductive reasoning, reasoning by analogy, cognitive biases in scientific reasoning), but they are usually specifically interested in descriptive, more empirical aspects of scientific reasoning. Cognitive philosophers of science approach the issue in a more theoretic and often normative way. Cognitive philosophy of science may, for instance, include analyses of the norms of scientific reasoning and models of reasoning that may try to make sense of the descriptive accuracy (e.g., by looking at how epistemic norms fit with our cognitive tasks in science, or by inspecting various accounts of scientific inference).

It remains an open question whether there is a clear demarcation between cognitive science of scientific reasoning and cognitive philosophy of science – a question that perhaps needs not be resolved as long as cognitive philosophy of science leads to novel and fruitful insights. And indeed, it seems to do just this.

In our symposium we represent and discuss a lively variety of topics in cognitive philosophy of science with the aim of showing how cognitive science and philosophy of science intertwine and inform each other in a number of interesting ways. We want to explore the following guiding questions:

1) What are some of the topics addressed by cognitive philosophy of science?
2) How can and how does cognitive science play a role in contemporary philosophy of science?
3) Can cognitive philosophy of science also inform cognitive science, and if so, by what means?
4) What is specific of the interaction between cognitive science and philosophy of science?

Overview of the talks:

Matteo De Benedetto will open the symposium with a talk about how the discussion about theoretical terms in science can be informed by models of ad hoc concepts in cognitive science. Nina Poth will analyze recent discussions on unification in cognitive science in the light of debates on unification in general philosophy of science. Daniel Kostić will then offer an empirical interpretation of the feeling of understanding based on topological explanations of cognitive insight. Mel Andrews offers an account of the epistemic efficacy of mathematical modelling, illustrated with detailed case studies from cognitive science. Borut Trpin will conclude the symposium by discussing how probabilistic information may lead to changes in how a rational agent perceives the causal structure of a certain situation.

Significance

The symposium is dedicated to cognitive philosophy of science, a subfield of philosophy of science which is related but essentially distinct from philosophy of cognitive and behavioural sciences. The former denotes the mutual applications between cognitive science and philosophy of science, while the latter denotes the philosophy of a specific science. Although cognitive philosophy of science is a lively approach, we believe that it is not typically recognised as a specific line of research in philosophy of science. Hence, we believe that our symposium will provide information about an important if perhaps overlooked trend in philosophy of science. Moreover, the participants will address a number of topics that are interesting on their own, while also showing how cognitive modeling and other methods, standard for cognitive science, may further these specific discussions which are more often addressed in non-cognitive philosophy of science. Finally, we want to show how cognitive philosophy of science proceeds in practice to thereby make the use of such methods more salient. We also expect a lively discussion about ways in which this subfield of philosophy of science may further develop.

Matteo De Benedetto (Ruhr-Universität Bochum)

Theoretical Terms as Ad Hoc Concepts

In this talk, I'll propose a novel perspective on the problem of theoretical terms in science, arguing that we should understand them as a specific type of ad hoc concepts, as the term has been used in psychology (Barsalou, 1983). Traditionally, theoretical terms have been understood in analogy with kind terms (e.g. Schwartz 1980), i.e. terms that express kinds concepts and that (possibly) refer to kinds categories. Consistently, the cognitive relata of theoretical terms are usually considered to be noun/taxonomic concepts, i.e. context-independent concepts stored in our long-term memory. Even deflationary and anti-realist accounts of theoretical terms understand them as lesser ontological or epistemological versions of kind terms (e.g. Laporte 2003, Strevens 2012).

I will argue against this traditional assumption of theoretical terms as analogous to kind terms by proposing an alternative view that conceptualizes theoretical terms as analogous to ad hoc concepts (Barsalou, 1983). Ad hoc concepts are concepts such as 'things to take in case of a fire' or 'clothes to bring on a ski holiday'. These concepts are, in contrast to noun/taxonomic concepts, highly contextual concepts, created on the fly in relation to a specific cognitive goal. I will argue that many of our best semantics for theoretical terms in science stress their open-endedness (Carnap, 1956), scale-sensitivity (Wilson 2017, Bursten 2018), and context-dependency (Cartwright 1983, Wilson 2006). I will then show how these properties are typical features of terms related to ad hoc concepts. Building on these semantic analogies, I will argue that theoretical terms are better characterized, from both a semantical and an epistemological point of view, as ad hoc concepts. Finally, I will sketch some implications of my proposal for general debates in philosophy of science closely connected to the problem of theoretical terms, such as the discussion on scientific realism, the debate about scientific kinds, and the problem of incommensurability.

References:

- Barsalou, L.W. (1983): "Ad Hoc Categories". Memory & Cognition 11, 211-227.
- Bursten, J.R. (2018): "Smaller than a Breadbox: Scale and Natural Kinds". British Journal for Philosophy of Science 69 (1), 1-23.
- Carnap, R. (1956): "The Methodological Character of Theoretical Concepts". In Feigl, H. and Scriven, M. (Eds.), The Foundations of Science and the Concepts of Psychology and Psychoanalysis, University of Minnesota Press, Minneapolis, 38-76.
- Cartwright, N. (1983): How the Laws of Physics Lie. Oxford University Press, Oxford.
- LaPorte, Joseph (2003). *Natural Kinds and Conceptual Change*. Cambridge University Press.
- Schwartz, S.P. (1980): "Natural Kind Terms". Cognition 7, 301-315.
- Strevens, Michael (2012). Theoretical terms without analytic truths. *Philosophical Studies* 160 (1):167-190.
- Wilson, M. (2006): Wandering Significance: An Essay on Conceptual Behavior. Clarendon Press, Oxford.
- Wilson, M. (2017): Physics Avoidance: And Other Essays in Conceptual Strategy. Oxford University Press, Oxford.

Dimensions of Unification in Cognitive Science

Nina Poth (Ruhr-Universität Bochum)

What makes for a good account of cognition? When answering this question, philosophers of cognitive science sometimes appeal to strategies of unification. *Simplicity*, *unbounded scope*, and *beauty* have become especially popular criteria to evaluate unification strategies in cognitive science (Milkowski 2016, Milkowski & Hohol 2021). I argue that the common interpretations of these criteria lack general plausibility since they apply only in the narrow domain of mechanistic approaches to cognition. For instance, these interpretations do not appropriately capture the common virtues associated with typical cases of unifying accounts in Bayesian cognitive science. The problem posed is that there are currently at least two disconnected sets of dimensions of unification, and so there is no unitary basis to jointly evaluate unifications from different areas in cognitive science. From the perspective of mechanistic approaches, unifying accounts of cognition are appropriate to the extent that they are ontologically simple or parsimonious, non-monstrous, and invariant (ibid.). From the perspective of Bayesian approaches, unifying accounts of cognition are appropriate to the extent that they are formally elegant and unbounded in scope (Colombo & Hartmann 2017). However, an account can be parsimonious and monstrous without being elegant, and invariant without being of a broad scope (or vice versa); so it is not clear how these two different sets of dimensions can be aligned within the field. To remedy this correspondence problem, I propose replacing the popular dimensions with a criterion of *mutual informational relevance*, which is inspired by Myrvold's (2003, 2017) approach to unification in the general philosophy of science. Based on two case studies, one on Bayesian models of concept learning and the other on Predictive Processing models of cognition, I show that this dimension obtains a single coherent interpretation that can be applied to both mechanistic and Bayesian models in cognitive science alike.

References

Colombo, M., & Hartmann, S. (2017). Bayesian Cognitive Science, Unification, and Explanation. *British Journal for the Philosophy of Science, 68(2),* 451-484.

Miłkowski, M., & Hohol, M. (2021). Explanations in cognitive science: unification versus pluralism. *Synthese*, *199(1),* 1-17.

Miłkowski, M. (2016). Unification strategies in cognitive science. *Studies in Logic, Grammar and Rhetoric*, *48(1),* 13-33.

Myrvold, W. C. (2003). A Bayesian account of the virtue of unification. *Philosophy of Science*, *70(2),* 399-423.

Myrvold, W. C. (2017). On the evidential import of unification. *Philosophy of Science*, *84(1),* 92-114.

Topological explanation of the sense of understanding

Daniel Kostić (Radboud University)

Philosophers have debated whether “Aha!” moments amount to mere *feelings* of understanding or whether they reliably track with the genuine *achievement* of understanding a phenomenon (Regt 2004; Grimm 2010; Trout 2005; 2007). However, little philosophical work has examined how the cognitive processes that support these moments can provide useful explanatory heuristics.

I argue that the psychological phenomenon of cognitive insight, has the potential to shed light on philosophical debates about the feeling of understanding. I discuss a case study by Schilling (2005), which uses network models to explain the phenomenon of “cognitive insight,” the discovery of a novel solution to a problem. In her model, nodes are representations of problems and edges are associations between these representations. When such networks have a small-world topology, this means that it will be easier to associate the representation of one problem with another. The small-world topology is characterized by short path lengths (average number of edges that need to be traversed to reach any node in a network), and high clustering coefficient (the degree of interconnectedness of nodes within the same neighborhood).

Schilling takes this to be the core of cognitive insight showing how various processes that precipitate cognitive insight—"(a) completing a schema, (b) reorganizing visual information, (c) overcoming a mental block, (d) finding a problem analog, and (e) random recombination”—change the topological structure so as to either decrease average path length or increase the network’s clustering coefficient (Schilling, 2005, p. 134), and in that way showing how the cognitive insight counterfactually depends on topological properties.

Schilling’s model hence provides a distinctively topological explanation (Kostić 2020) of cognitive insight, which shows that “Aha!” moments can be genuine instances of understanding as long as they are underlined by appropriate network of contextual information which is organized as a small-world topology.

References:

Grimm, Stephen R. 2010. “The Goal of Explanation.” *Studies in History and Philosophy of Science Part A* 41 (4): 337–44. https://doi.org/10.1016/j.shpsa.2010.10.006.

Kostić, Daniel. 2020. “General Theory of Topological Explanations and Explanatory Asymmetry.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 375 (20190314): 1–8. http://dx.doi.org/10.1098/rstb.2019.0321.

Regt, Henk W. De. 2004. “Discussion Note: Making Sense of Understanding.” *Philosophy of Science* 71 (1): 98–109. https://doi.org/10.1086/381415.

Schilling, Melissa A. 2005. “A ‘Small-World’ Network Model of Cognitive Insight.” *Creativity Research Journal* 17 (2–3): 131–54. https://doi.org/10.1080/10400419.2005.9651475.

Trout, J D. 2007. “The Psychology of Scientific Explanation.” *Philosophy Compass* 2/3: 564–91. https://doi.org/10.1111/j.1747-9991.2007.00081.x.

Trout, J. D. 2005. “Paying the Price for a Theory of Explanation: De Regt’s Discussion of Trout*.” *Philosophy of Science* 72 (1): 198–208. https://doi.org/10.1086/426849.

Between Oracular Modelling & the Theory-Free Ideal

Mel Andrews (University of Cincinnati)

Certain approaches to mathematical modelling and quantitative analysis in the sciences are more effective than others. As background, I sketch an account of the epistemic aims of science and the nature and efficacy of mathematical modelling. I then look to the cognitive and neurosciences, drawing out a characterisation of two epistemically impotent strategies of mathematical modelling and quantitative analysis and one highly effective strategy. The impotent methods include, in the first case, the use of off-the-shelf techniques for quantitative analysis that leave the formal models they rest on implicit, thus baking-in unwarranted assumptions and paving the way for deceptive interpretative practices. I take this first approach to exemplify what I have termed the *theory-free ideal*. In the second case, we observe the construction, elaboration, and deployment of models that lack sufficient contact with data. This I dub *oracular modelling*. We can see the first approach play out in the common practice of *p*-hacking, resulting in an abundance of experimental paradigms that have failed to replicate across the psychological and cognitive sciences; the second approach is exemplified by work on the FEP and IIT in theoretical neuroscience, where modelling frameworks are treated as though capable of revealing knowledge about nature without making appropriate contact with data. Effective mathematical modelling strategies neither presume that data will reveal the causal structure of the world without the careful deployment of formal models nor look to formal models to inform us about the world without connecting them up to the results of a measurement procedure. Instead, they facilitate a dialogue with data. Exemplified by early work in psychophysics and more contemporary work in mathematical psychology, effective mathematical modelling strategies build formal models purpose-fit to answering specific empirical questions about specific cognitive phenomena. This requires, however, a high degree of experimental and mathematical literacy and a willingness to engage in theorising.

Allen, C. (2014). Models, mechanisms, and animal minds. *The Southern Journal of Philosophy, 52*, 75-97.

Andrews, M. (2021). The math is not the territory: navigating the free energy principle. *Biology & Philosophy, 36*(3), 1-19.

Andrews, M. (2022). Making Reification Concrete: A Response to Bruineberg et al. *Brain & Behavioural Sciences.*

Machery, E. (2021). The alpha war. *Review of Philosophy and Psychology, 12*(1), 75-99.

Potochnik, A. (2017). *Idealization and the Aims of Science*. University of Chicago Press.

Borut Trpin (MCMP/LMU Munich)

Learning Causal Structure: A Bayesian Approach

In this talk, which is based on joint work with Ulrike Hahn (Birkbeck London) and Stephan Hartmann (LMU Munich), I will address how probabilistic or correlational information may lead to changes in perceived causal structure.

There are many things we believe probabilistically. Some beliefs about these probabilities imply particular causal structures. For instance, if we already believe that a mobile phone's battery is about to run out soon, then seeing a low battery warning has no additional effect on our belief that we will not be able to hold a long telephone conversation without charging the phone. More formally, our probabilistic beliefs imply that a low battery warning and the duration of immediate calls are conditionally independent given a dying battery. On the other hand, learning that there probably is no cell reception has an impact on immediate phone calls -- the effect of this cause is that there will likely be no calls. Some learning experiences therefore lead to changes in causal understanding, while others do not.

Given that our probabilistic beliefs seem to imply specific causal structure, we can then consider an interesting question: how, if at all, should the perceived causal structure change after learning new evidence? Under which conditions do we expect that certain or uncertain evidence might lead to learning causal structure?

I will introduce a recent Bayesian approach to modelling rational belief change that allows us to incorporate much more general types of evidence than standard Bayesian conditionalisation -- the so-called distance based approach to Bayesianism (DIST-Bayes). The approach, roughly, says that a rational belief update may best be seen as a statistical distance minimization problem. I will inspect some numeric examples and formal results that point out when belief change might lead to changed understanding of causal structure. Notably, by means of modelling we can find situations in which some distance measures will predict that the causal structure should be retained after incorporating new evidence, while the others will not.

Matteo De Benedetto

Since October 2021, I am a postdoctoral researcher at the Ruhr University Bochum in the Emmy-Noether research group “From Perception to Belief and Back Again”. I recently defended my PhD at the Munich Center for Mathematical Philosophy. (MCMP, 2017-2022).

My research focuses on the phenomenon of conceptual change, broadly understood as the many philosophically interesting ways in which our concepts change. My work engages with several philosophical areas, combining traditional philosophical and historical methodologies with formal tools from logic, mathematics, and cognitive science. I am particularly interested in Carnapian explication, scientific change, scientific theory choice, cognitive models of science, conceptual inferences, belief revision, and the Church-Turing Thesis.

Nina Poth

Nina Poth is currently a Postdoctoral Researcher at the Institute of Philosophy II at Ruhr University Bochum, where she is researching on the philosophical foundations of Bayesian models of cognition. She obtained her PhD in Philosophy from the University of Edinburgh in 2020 with a thesis on perceptual categorisation, Bayesian inference and psychological similarity. Before that, she studied in the MSc Cognitive Science and BA Philosophy & Social Science programs at Ruhr University Bochum, with a research visit in the Department of Philosophy (KGW) at the University of Salzburg (2016).

Nina's main research interests are in philosophy of mind and cognition, epistemology, and philosophy of science. Within each of these areas, she is especially interested in questions about computational models of cognition, concept learning, rationality, and issues of unification and explanation in cognitive science.

DR. DANIEL KOSTIĆ

Institute for Science in Society (ISiS), Radboud University, Huygens Building, Heyendaalseweg 135
6525 AJ, Nijmegen, The Netherlands

Phone: +31 (0) 6 25 08 70 01
Email: daniel.kostic@gmail.com
Web: https://daniel-kostic.weebly.com

Areas of Specialization: Philosophy of science, philosophy of neuroscience, philosophy of mind.
Areas of Competence: Philosophy of ecology, metaphysics.

CURRENT APPOINTMENT

2020-2022 Radboud Excellence Initiative Fellow, Institute for Science in Society (ISiS) Radboud University.

EDUCATION

2012 Humboldt Universität zu Berlin, PhD in Philosophy.
Thesis Title "The explanatory gap problem: how neuroscience might contribute to its solution".

SELECTED PUBLICATIONS

1. Kostic, D. and Khalifa, K. (Forthcoming). "Decoupling Topological Explanations from Mechanisms." *Philosophy of Science*.
2. Kostić, D. and Khalifa, K. (2021) "The Directionality of Topological Explanations." *Synthese,* https://doi.org/10.1007/s11229-021-03414-y.
3. Kostić, D. (2020). "General Theory of Topological Explanations and Explanatory Asymmetry". *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375: 20190321.
4. Kostić, D. (2019). "Minimal Structure Explanations, Scientific Understanding and Explanatory Depth", *Perspectives on Science*, 27 (1), 48-67.
5. Kostić, D. (2018). "The Topological Realization". *Synthese*, 195(1), 79-98.

SELECTION OF GRANTS

2020-2022 Radboud Excellence Initiative Fellowship (Radboud Excellence Initiative). *Total amount: 205.300 €.*

2016-2018 Marie Skłodowska-Curie Fellowship (EU's H2020-MSCA-IF-2015 Programme). *Total amount: 185.076 €.*

SERVICE TO THE PROFESSION

2020 Co-Founder of The Dutch Distinguished Lecture Series in Philosophy and Neuroscience. http://daniel-kostic.weebly.com/dutch-distinguished-lecture-series.html

2019 Co-founder of the Annual Workshop Series "Scientific Understanding and Representation (SURe). https://sure-workshop.weebly.com

2015 Co-founder and coordinator (until 2020) of the East European Network for Philosophy of Science (EENPS). https://eenps.weebly.com

Mel Andrews

Mel Andrews is a current doctoral student and instructor in the department of philosophy at the University of Cincinnati and a Principles of Intelligent Behavior in Biological and Social Systems fellow. In 2021 Mel was a visiting scholar at the University of Pittsburgh department of History & Philosophy of Science. They completed their undergraduate degree in psychology, cognitive & brain sciences from Tufts University (2018). Mel's research has focused on machine learning, mathematical modelling in science, cognitive science, and general philosophy of science.

Borut Trpin

Borut Trpin obtained a PhD degree (2018) in philosophy at the University of Ljubljana with a dissertation on learning from conditionals. In 2018 he was awarded an Ernst Mach Grant for postdoctoral research, which he conducted in the Department of Philosophy (KGW) at the University of Salzburg (study year 2018/2019). In 2019 he was awarded a Humboldt Research Fellowship for Postdoctoral Researchers to research probabilistic versions of inference to the best explanation at the Munich Center for Mathematical Philosophy (MCMP, 2019-2021). He has been a postdoctoral researcher in a DFG project on descriptive and normative reasoning at the MCMP since 2021, and since 2022 also a researcher in a project on thought experiments at the University of Maribor.

Borut's main research interests are in epistemology, philosophy of science, and reasoning broadly conceived. Some of the topics that particularly interest him are inference to the best explanation, models of human reasoning, coherence, and scientific disagreements. He likes to use computer simulations for philosophical research.

The EENPS 2022 Organizers

Program committee

Co-chairs:

- Sorin Bangu (University of Bergen)
- Vlasta Sikimić (University of Tübingen)

Members:

- Ana-Maria Cretu (Bristol)
- Angela Potochnik (Cincinnatti)
- Ave Mets (Tartu)
- Borut Trpin (LMU)
- Catherine Herfeld (Zurich)
- Chris Pincock (OSU)
- Christian Feldbacher-Escamilla (Cologne)
- Ciprian Jeler (Iasi)
- Edit Talpsepp (Tartu)
- Elena Trufanova (Institute of Philosophy, Russian Academy of Sciences)
- Iulian Toader (Vienna)
- Jan Sprenger (Turin)
- Jo Wolf (Edinburgh)
- Julie Jebeille (Bern)
- Julie Zahle (Bergen)
- Laura Franklin-Hall (NYU)
- Magdalena Małecka (Aarhus)
- Marcin Milkowski (Warsawa)

- Maria Serban (UEA)
- Marianna Antonutti-Marfori (Paris)
- Michael Baumgartner (Bergen)
- Miklos Redei (LSE)
- Nic Fillion (Simon Fraser)
- Patricia Palacios (Salzburg)
- Richard Dawid (Stockholm)
- Sabina Leonelli (Exeter)
- Stathis Psillos (Athens)
- Tudor Baetu (Trois Rivieres, QC, Canada)
- Vincent Ardourel (Paris)

Local organizing committee

- Endla Lõhkivi (Chair (to March 2022), University of Tartu)
- Jaana Eigi-Watkin (Chair (from March 2022), University of Tartu)
- Riin Kõiv (University of Tartu)
- Kristin Kokkov (University of Tartu)
- Eveli Neemre (University of Tartu)
- Katrin Velbaum (University of Tartu)
- Peeter Müürsepp (Tallinn University of Technology)

Acknowledgement

The EENPS 2022 Tartu conference is supported by the Estonian Research Council (ERC) grant no. 462 "Philosophical Analysis of Interdisciplinary Research" 2019-2023, Institute of Philosophy and Semiotics, University of Tartu, Faculty of Humanities and Arts, University of Tartu.

Author Index