BSPS British Society for the Philosophy of Science

Annual Conference

5-7 July 2023

University of Bristol

Abstracts for Contributed Papers

Formal Philosophy of Science

Entropic taming of the Look Elsewhere Effect

Miklós Rédei[1], Márton Gömöri[2]

[1]Department of Philosophy, Logic and Scientific Method, LSE, London, United Kingdom. [2]Department of Logic, Eötvös Loránd University, Budapest, Hungary

The Look Elsewhere Effect (LEE) is an important phenomenon in parameter dependent statistical hypothesis testing based on p-values [1]. The problem of how to mitigate it is especially acute in experimental high-energy particle physics [3]. The LEE is one of the reasons why in high-energy particle physics the significance level of rejecting a hypothesis is required to be very high. One can argue however, as Dawid [2] does, that the usual reasoning referring to the LEE to motivate the demand of high significance in particle physics can be too much decoupled from the theoretical background of the experiment. Dawid suggests a Bayesian way of taking into account the physicists' prior-to-test theoretical belief in the truth of the theory under test to mitigate the LEE.

Motivated by Dawid's analysis, in this paper we suggest another Bayesian way of mitigating the LEE. Our main idea is to adjust the significance level by taking into account a possible prior-to-test theoretical knowledge that indicates the probability with which the (dis)confirming phenomenon might be found at different parameter values in the test. We demand the adjustment to satisfy four requirements:

(i) The correction should be sensitive to the uncertainty embodied by the probability measure ρ representing the scientists' prior theoretical knowledge/belief about at which parameter value the confirming evidence might be found in a test: the higher the uncertainty of ρ, the higher significance is required.

(ii) The correction should also be sensitive to the prior probability $\rho(t)$ of the observed event to be found in the test at a specific parameter value t: the higher $\rho(t)$ the smaller significance is required.

(iii) The adjustment should yield the “Bonferroni correction” [1] in the case when ρ is the uniform probability measure on the parameter space.

(iv) When ρ is totally concentrated at a particular value of the parameter and hence represents maximal certainty, the correction should be zero: no need to adjust the p-value as a result of the LEE.

Taking the entropy of ρ as a measure of uncertainty of ρ, we define a quantitative entropic correction of significance that depends on the entropy of ρ, on the value $\rho(t)$ and on the number of parameters, and which satisfies the conditions (i)-(iv). Thus the proposed entropic correction mitigates the LEE in a way that reflects specific structural features of the prior knowledge represented by ρ.

The entropic taming of the LEE allows one to introduce in a disciplined and technically explicit manner a Bayesian component in the otherwise frequentist statistical theory testing. This is in the spirit of Dawid. But the kind of Bayesianism the entropic correction represents differs significantly from the sort of Bayesianism present in [2]. Dawid's analysis involves the subjective prior probabilities about the truth of the null hypothesis and of the theory to be tested. Such subjective priors are completely absent in the entropic correction of the significance levels: Degrees of beliefs enter the entropic correction of significance only through the prior-to-test probability ρ representing expectations about where the confirming evidence might occur. Also, Dawid's analysis assumes that, using our terminology, ρ is the uniform probability – and the main point in the entropic correction is that ρ need not be uniform.

[1] A.E. Bayer, U. Seljak. The look-elsewhere effect from a unified Bayesian and frequentist perspective. Journal of Cosmology and Astroparticle Physics, 2020(10):1–22.

[2] R. Dawid. Bayesian perspectives on the discovery of the Higgs particle. Synthese, 2017(194):377–394.

[3] L. Lyons. Bayes and Frequentism: a Particle Physicist’s perspective. Contemporary Physics, 2013(54):1–16.

What is Credit in Science? A Value-Based Interpretation of the Credit Maximisation Approach to the Social Philosophy of Science

Thijs Ringelberg

University of Groningen, Groningen, Netherlands

What is Credit in Science?

Questions concerning the social organisation of science are addressed increasingly often (and increasingly successfully) by means of what might be called the Credit Maximisation Approach (CMA). This approach employs computational techniques to model the behaviour of scientific communities on the assumption that scientists act in pursuit of “social credit”, and thus are faced by an incentive structure called the “credit economy”. But although the functioning of the credit economy is clear in the context of formal models, there is a disconnect between the formal models and the reality which they are meant to represent. The aim of this paper, therefore, is to establish how these models should be interpreted. I argue that the most plausible interpretation 1) casts the credit economy as dependent on and reflective of a pre-existing normative consensus; 2) reveals this normative consensus to be centred on a new type of value, and geared toward the common epistemic good; and 3) restricts the ways in which the CMA can be employed to give policy advice.

I argue that the credit economy functions through an interplay between scientists’ attitudes and those scientific institutions I term “praise-institutions” (such as authorship conventions, citations, and academic prizes). Provided that this interplay works well, it is possible to be motivated in one's scientific work by credit incentives: a researcher might strive to make a discovery because she knows that to be the way to further her career. But due to the nature of the attitudes on which these incentives are based, the credit economy must reflect a pre-existing normative consensus, and cannot function autonomously.

The relevant attitudes are attitudes of esteem: scientists' evaluations of each other as scientists. The normative consensus that is a precondition for the functioning of the credit economy therefore focuses on the (shared) concept of a good scientist. I argue that standard analyses employing the CMA show us that this normative consensus is geared toward the common epistemic good: the fact that scientists endorse this picture of a good scientist, and not some other picture, has positive epistemic consequences for the scientific project.

This value-based understanding of the CMA has consequences for the way we can use the CMA to make science policy recommendations. Of central importance in this context are the praise-institutions, which function to communicate esteem-judgements across the scientific community. In order for this communication of esteem to be successful, the praise-institutions need to be perceived as tracking the normative consensus. Policy interventions which make adjustments to the praise-institutions run the risk of tarnishing this perception, thereby endangering the institution's capacity to communicate esteem. To establish whether such tarnishing will take place in a specific case, empirical knowledge is required about the normative consensus in the relevant scientific subdomain. Thus if we wish to base policy on the CMA, sociologists and social psychologists must be enlisted to explore the functioning of the credit economy in specific situations, lest the recommended policy interventions have unexpected effects.

References

Bright, Liam Kofi. 2021. "Why Do Scientists Lie?" Royal Institute of Philosophy Supplement 89 (May): 117–29. https://doi.org/10.1017/S1358246121000102.

Heesen, Remco. 2017. "Communism and the Incentive to Share in Science." Philosophy of Science 84 (4): 698–716. https://doi.org/10.1086/693875.

Heesen, Remco, and Liam Kofi Bright. 2021. "Is Peer Review a Good Idea?" The British Journal for the Philosophy of Science 72 (3): 635–63. https://doi.org/10.1093/bjps/axz029.

Kitcher, Philip. 1990. "The Division of Cognitive Labor." The Journal of Philosophy 87 (1): 5. https://doi.org/10.2307/2026796.

Strevens, Michael. 2003. "The Role of the Priority Rule in Science:" Journal of Philosophy 100 (2): 55–79. https://doi.org/10.5840/jphil2003100224.

Ontic Structural Realism and Accounts of Theories

Teodor-Tiberiu Calinoiu

Lingnan University, Hong Kong, Hong Kong

The semantic account holds that scientific theories are collections of models (plus a theoretical hypothesis). Hans Halvorson (2012) formulated a couple of objections against language-free articulations of the semantic account, where models are purely non-linguistic entities: such articulations provide neither necessary, nor sufficient identity conditions for theories. He concludes that non-linguistic versions of the semantic account are untenable (and that the dispute between syntacticists and semanticists is spurious). Halvorson extends his argument to an objection against ontic structural realism (OSR), a position in the metaphysics of science holding (roughly) that all there is in reality are irreducible structures (Ladyman 1998, Ladyman and Ross 2007, French 2014). Halvorson argues that OSR is dependent upon a (language-free) semantic account, and, since the semantic account fails, OSR follows in its footsteps. His argument can be summarised as follows:

1. If OSR is true, then (a version of) the language-free semantic account is true.
2. The language-free semantic account is false.
3. Hence, OSR is false.

Halvorson argues extensively for 2. However, his ground for 1 is purely exegetical, quoting the testimony of prominent structuralists (Ladyman 1998, Ladyman and French 2003). Granting that the semantic account misrepresents scientific theories (premise 2), I focus on the relationship between OSR and conceptions of scientific theories.

I argue that premise 1 fails. I reconstruct three arguments which attempt to establish OSR's dependence upon a language-free version of the semantic account, and I argue that all three fail either due to a commitment to an overly narrow space of possibilities regarding conceptions of theories, or to an unsanctioned inflation of the relationship between OSR's physical structures and the semantic account's models constituting the content of scientific theories. OSR comes in several versions, as does the semantic account: I shortly discuss the most prominent ones and their relations, while focusing on Ladyman and Ross's 2007 and French 2014's versions of OSR, and their preferred articulation of the semantic account, the partial structures approach. I acknowledge the motivation behind OSR's perceived reliance upon the semantic account and sketch an error-theory: if OSR is true, then, on good naturalistic grounds, scientific theories are ontologically committed to irreducible structures in reality, which may plausibly and more readily be represented by models furnished by the semantic account. However, I show that scientific representation by purely non-linguistic models is neither the sole option, nor the best. Ontological commitment to irreducible structures in reality doesn't transfer commitment to identifying scientific theories with collections of extra-linguistic entities. I conclude shortly addressing the reverse relation, sketching conditions under which a language-free version of the semantic account commits to a version of OSR.

Bibliography

1. Halvorson, H. (2012). What Scientific Theories Could Not Be. Philosophy of Science, 79(2), 183-206. doi:10.1086/664745
2. French, S. (2020). There are no such things as theories. Oxford University Press.
3. Frigg, R. (2022). Models and theories: A philosophical inquiry. Taylor & Francis.
4. French, S., & Saatsi, J. (2006). Realism about structure: The semantic view and nonlinguistic representations. Philosophy of Science, 73(5), 548-559.
5. Ladyman, J. (1998). What is structural realism?. Studies in History and Philosophy of Science Part A, 29(3), 409-424.

General Philosophy of Science

Sisyphean Science: Why Value Freedom Is Worth Pursuing

Menon Tarun[1], Jacob Stegenga[2]

[1]Azim Premji University, Bangalore, India. [2]Cambridge, Cambridge, United Kingdom

The ambition of this paper is to articulate a novel version of the value-free ideal which avoids the existing philosophical challenges. Our core argument is that the feasibility and desirability of attaining an end can be decoupled from the feasibility and desirability of pursuing that end, as illustrated in the example from the Indian philosophical tradition. A particular end-state may be unfeasible or undesirable, yet pursuing that end may nevertheless be feasible or desirable. The ideal of world peace may be impossible to achieve, yet pursuit of that ideal is both possible and good; de-escalating conflicts, disarmament, and a more equitable distribution of resources are all potential means to approach world peace, and these means are, at least to some degree, feasible to enact. Things are a little trickier when we consider desirability. It may seem absurd to suggest that it is desirable to pursue an end that one judges undesirable, but some philosophers take this possibility seriously. Challenges to the value-free ideal have focused on the feasibility and desirability of a value-free end-state. Yet, one can grant that the value-free ideal is neither end-state feasible nor end-state desirable, while maintaining that pursuit of the end-state is both feasible and desirable. That is our goal.

Our primary aim is to argue that pursuit of the value-free ideal is desirable even if the end-state is undesirable. The conclusion of our argument is a specific—and as far as we know, novel—version of the value-free ideal, which holds that scientists ought to act as if science should be value-free.

The value-free ideal holds that non-epistemic values ought not influence the internal features of scientific reasoning. “Influence”, however, is a broad notion. The version of the ideal we defend opposes values playing a crucial role in the inference from evidence to conclusions, in the following sense: if an appeal to value V1 appears in the inferential chain connecting evidence E to conclusion C1, then replacing V1 with an alternate value V2 should not lead to a conclusion C2 that is inconsistent with C1. In other words, given a fixed evidential basis, whether a researcher regards a conclusion as true or false should not depend on the values they endorse. A scientific inference that is not value-free will have bifurcation points, places in the inferential chain where choosing between two value sets will lead to incompatible conclusions. The value-free ideal, as we conceive it, holds that science should aim at the elimination of all bifurcation points, and progress towards the ideal can be made by reducing the number of bifurcation points.

We proceed to describe various routine scientific strategies for eliminating bifurcation points, such as increasing the quality of evidence or hedging conclusions. These strategies are not cost-free, however, and various constraints on science, particularly policy-oriented science, limit the extent to which these strategies can be applied.

In short, we argue that it is possible to pursue the value-free ideal, it is good to pursue the value-free ideal, and we give the value-free ideal a novel articulation: scientists should act as if science should be value-free.

Alexandrova, Anna. 2018. “Can the Science of Well-Being Be Objective?” British Journal for the Philosophy of Science 69 (2): 421-445.

Douglas, Heather. 2000. “Inductive Risk and Values in Science.” Philosophy of Science 67 (4): 559-579.

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

Frisch, Mathias. 2020. “Uncertainties, Values, and Climate Targets.” Philosophy of Science 87 (5): 979-990.

Stating Structural Realism: Mathematics-First Approaches to Physics and Metaphysics.

David Wallace

University of Pittsburgh, Pittsburgh, USA

Structural realism rests on a seductive idea: that the descriptive categories of the scientific realist and the analytic metaphysician draw distinctions too fine for the scientist. Theory change might involve a radical shift of ontology, from waves as disturbances in aether to waves as self-subsistent field states or from heat as invisible fluid to heat as disordered motion, but the equations --- the structure --- of the theory change more continuously or not at all, and so knowledge of structure can be robust against theory change. Considerations of ontology might suggest a multiplying of theories, where gravity can be understood as curvature or as universal force, and where fields can be understood as extended matter or as properties of spacetime, but the equations don't care, and so underdetermination does not threaten knowledge of structure.

In the epistemic form of structural realism (ESR), this is a limitation on what we can know: science only tells us the structural features of reality, so any non-structural features, if knowable at all, are not knowable through science. In ontic structural realism (OSR), it transforms into a radical metaphysical thesis: if science need structure, and if metaphysics should conform toscience, then our metaphysics should contain nothing but structure. Both forms of structural realism --- but especially the ontic variety --- appear to challenge, even threaten, `traditional' analytic metaphysics, which seems committed to more than the structure that science reveals.

But throughout this vibrant literature it has been frustratingly difficult to pin down just what structural realism (in either form) actually says, and attempts to precisify it seem to lead away from the original seductive idea and back to supposedly-unscientific analytic philosophy: precise versions of ESR struggle to articulate a notion of structure intermediate between full-fat scientific realism on the one hand and instrumentalism on the other; precise versions of OSR get into surprisingly metaphysical conversations about how we can have `relations without relata', whether structural realism requires an infinite tower of individuals at one level reimagined as bundles of relations at another, and the like.

The purpose of this paper is to define a version of structural realism that realizes the core goals structural realists express and that is at least explicable to their metaphysical critics. The approach I propose is based on what I call a `maths-first' approach to physical theories (close to the so-called `semantic view of theories') where the content of a physical theory is understood primarily in terms of its mathematical structure and the representational relations it bears to physical systems, rather than as a collection of sentences that attempt to make true claims about those systems (a `language-first' approach). I argue that adopting the maths-first approach already amounts to a form of structural realism, and that the choice between epistemic and ontic versions of structural realism is then a choice between language-first and maths-first views of metaphysics. I then explore the status of objects (and properties and relations) in fundamental and non-fundamental physics for both versions of maths-first structural realism.

Indicative Bibliography

Dennett, D.C. (1991). "Real patterns." J.Phil 87, 27-51.

French, S. and J.Ladyman (2003). "Remodelling structural realism: Quantum physics and the metaphysics of structure." Synthese 136, 31-56.

Ladyman, J. and D.Ross (2007). Every Thing Must Go: Metaphysics Naturalized. Oxford: Oxford University Press.

McKenzie, K. (2014). "Priority and particle physics: Ontic structural realism as a fundamentality thesis." Brit.J.Phil.Sci. 65, 353-380.

Sider, T. (2020). The Tools of Metaphysics and the Metaphysics of Science. Oxford: Oxford University Press.

Elucidating and Embedding: Two Functions of How-Possibly Explanations

Franziska Reinhard

University of Vienna, Vienna, Austria

Scientific explanations do not always refer to how a phenomenon actually happened, but often to how it could have happened – the involve possibilities. Realizing this, philosophers of science have variously tried to characterize how-possibly explanations (HPEs) and distinguish them from how-actually explanations

(HAEs). Most of this discussion has focused on whether HPEs and HAEs are independent types of explanation (e.g. Dray, 1957; Forber, 2010), or whether they are of the same type but vary in their degrees of empirical support (e.g. Brandon, 1990; Bokulich, 2014). In this paper, I argue that the existing debate over the nature of HPEs has failed to pay adequate attention to the different, but complementary, *functions* possibilities play in scientific investigations and explanations. To bring these functions to the fore, I introduce the following distinction:

Elucidating HPEs: Specifying and demonstrating a possible process accounting for a research target.

Embedding HPEs: Demonstrating how a research target fits into a space of suitably constrained possibilities.

I specify both functions of HPEs with reference to two scientific case studies from prebiotic chemistry, a subfield of origins-of-life research. Prebiotic chemists try to understand how biomolecules formed from simple precursors on the Earth more than 3.5 billion years ago. This is a promising case for exploring the role of possibilities in science because investigating the origins-of-life is so difficult. Due to a lack of direct evidence for early Earth processes, researchers can often only try to infer possible, or perhaps plausible, explanations for how biomolecules and ultimately life first emerged. As I will highlight, there are at least two strategies to do so, corresponding to my distinction between elucidating and embedding HPEs. The first strategy involves developing possible pathways for biomolecules under laboratory conditions approximating the early Earth environment, thereby ultimately providing what I call an elucidating HPE. The second strategy tries to uncover salient properties of existing biomolecules in reference to chemical alternatives that could have formed on the early Earth under similar conditions, which amounts to providing an embedding HPE.

I contrast my distinction between elucidating and embedding HPEs with a recent proposal by Wirling and Grüne-Yanoff (forthcoming). According to them, the debate over the nature of HPEs can be reconciled by paying attention to the different ***types*** of possibilities presupposed in the discussion – some accounts of HPEs presuppose objective possibilities (possibilities for how the world could be), while others presuppose epistemic possibilities (possibilities compatible with our knowledge of the world). I argue that, while Wirling and Grüne-Yanoff are right to point out that HPEs come in different forms, their proposal ultimately misses important aspects of the scientific practices in my case studies. As I point out, in practice, considerations about objective and epistemic possibilities are hard to keep apart. Finally, I spell out the consequences of my distinction between elucidating and embedding HPEs for the debate over the nature of HPEs.

Bokulich (2014). How the Tiger Bush Got Its Stripes: 'How Possibly' vs. 'How Actually' Model Explanations. The Monist 97 (3).

Brandon (1990). Adaptation and Environment. Princeton University Press.

Dray (1957). Laws and Explanation in History. Greenwood Press.

Forber (2008). Confirmation and explaining how possible. Studies in History and Philosophy of Biological and Biomedical Sciences 41 (1).

Wirling & Grüne-Yanoff (forthcoming). Epistemic and Objective Possibility in Science. British Journal for the Philosophy of Science.

The Tracking View of Mathematical Explanation

Tuomas Tahko

University of Bristol, Bristol, United Kingdom

Is science-driven mathematical explanation explanatory in its own right or does it derive its explanatory power from some other relation or relations? This is a key question in the literature on mathematical explanation, but one that is difficult to assess without some appeal to a more general account of explanation. In the literature on metaphysical explanation (often used synonymously with 'grounding'), an analogous question has received some attention. Does grounding derive its explanatory power from a distinct form of determination or perhaps a variety of different kinds of metaphysical dependence relations? In other words, how can we distinguish between the epistemic, explanatory part of metaphysical explanation and the metaphysical or 'worldly' content – the metaphysical determination or dependence – that backs this explanatory part? The tracking view of metaphysical explanation suggests that grounding itself is not a form of explanation, but it does track either worldly grounding relations or other metaphysical dependence relations.

I will examine an analogous view of mathematical explanation and the dependence relation or relations that it could be considered to track. On the tracking view of mathematical explanation, the debate about the status of science-driven mathematical explanation turns on the nature of these dependence relations, namely, does mathematical explanation track a distinct form of mathematical dependence? While I will ultimately remain neutral on this issue, the upshot of the paper is that a close comparison of these two analogous debates is warranted.

The goals of the paper are as follows:

(1) To briefly illustrate a general view regarding the relationship between the epistemic ('explanatory') and the metaphysical ('worldly') nature of explanation.

(2) To show that this view of explanation can be applied to both metaphysical explanation and mathematical explanation, which leads to analogous questions about the relationship of the 'worldly' ontic dependence or determination and the explanation that it 'backs' in both debates.

(3) To suggest that a closer dialogue between these two debates is warranted, and has potential to be fruitful.

The focus is on mathematical explanation in science or 'science-driven mathematical explanation', as Baker (2012) puts it.

We can define the options succinctly. Does mathematical explanation track some distinctly mathematical dependence relation in the world, or does it track the same dependence relation or relations as other sciences? If mathematical explanation derives its explanatory power from the same kind of dependence that backs scientific explanations more generally, then it would appear that this type of mathematical explanation is just one of several epistemic aspects that we may associate with a broad class of dependence relations (cf. Saatsi 2011). Moreover, this might constitute a challenge to the explanatory indispensability argument.

But what if it turns out that mathematical explanation tracks a specific type of relation of dependence, which cannot be accounted for in terms of other dependence relations? Well, then it would appear that mathematical explanation is distinct from other familiar types of explanation, and this would perhaps be grist to the mill of the explanatory indispensability argument.

I will discuss the implications of each view, and their link to the literature on metaphysical explanation.

References

Baker, A. 2012. 'Science-Driven Mathematical Explanation.' Mind 121 (482): 243–267.

Lange, M. 2016. Because Without Cause: Non-Causal Explanations in Science and Mathematics. Oxford University Press.

Mancosu, P. 2008. 'Mathematical Explanation: Why it Matters.' In P. Manoscu (ed.), The Philosophy of Mathematical Practice. Oxford University Press, pp. 134–150.

Saatsi, J. 2011. 'The Enhanced Indispensability Argument: Representational versus Explanatory Role of Mathematics in Science.' British Journal for the Philosophy of Science 62: 143–154.

Steiner, M. 1978. 'Mathematical Explanation.' Philosophical Studies 34: 135–151.

Expressivism about Laws and Meta-laws

Josh Hunt

MIT, Cambridge, USA

Humean best systems accounts face a dilemma: in order to vindicate scientists' ordinary nomological discourse, they must embroil themselves in substantial metaphysics. To those averse to substantial metaphysical posits, this is a significant cost. Here, I develop an expressivist account of laws of nature that avoids this dilemma. My account simultaneously (i) vindicates scientists' ordinary law-claims while (ii) avoiding substantial metaphysical commitments. I argue that expressivism has distinct advantages over pragmatic Best Systems Accounts, such as those defended recently by Hicks (2018), Dorst (2019), and Jaag and Loew (2020). I thereby disagree with Callender's (2023) assessment that these two broadly Humean approaches are on a par.

To vindicate scientists' nomological discourse, I preserve the conceptual structure of Lange's (2009) account. Lange's account elegantly vindicates the intricate relationships between laws and meta-laws, such as when a space-time symmetry constrains the form of possible force laws. On a realist construal, Lange's account requires commitment to primitive counterfactuals. To avoid this commitment, I propose an expressivist account of subjunctive and counterfactual conditionals. To judge that a subjunctive is true is to express an attitude of ***being for planning for the consequent to be the case, in the event you learn that the antecedent has occurred***. Following Ward (2002), one can think of this as involving an attitude of ***being for inferring the consequent, in the event you learn the antecedent***. I treat counterfactuals as subjunctives involving past tense, counter-to-fact antecedents.

Substituting my account of counterfactuals into Lange's account leads to expressivism about laws of nature. To judge that a claim is a law is to express an attitude of ***being for planning for that claim to remain true in the event that any logically-consistent counterfactual antecedent has obtained***. In this way, scientists' nomological claims express attitudes about how best to plan for solving problems. Scientific practice crucially involves considering not just what has or will occurred, but also what might or might have occurred. Scientists arrive at their law claims by testing which plans work best for saving the phenomena, including predicting future phenomena.

By working within Lange's framework, my account improves on Ward's (2002) projectivist account of laws of nature. My account straightforwardly accommodates graded modal claims to the effect that some laws or symmetry principles are more necessary than others. This provides an advantage over Best Systems Accounts as well, which struggle to capture graded modality in a uniform and sufficiently general manner. My account also has a key advantage compared to recent pragmatic BSAs. These accounts problematically build the aims of science into the ***semantics*** for law-claims, making claims about laws mind-dependent. In contrast, my account respects the ordinary scientific conception of laws of nature as being mind-independent and objective: laws of nature do not depend on us or our aims. Rather, the aims of science figure only in the ***meta-semantics***, describing how agents like us have arrived at our concept of laws of nature and how we might improve this concept in the future.

Callender (2023). "Humean Laws of Nature: The End of the Good Old Days". In: ***Humean Laws for Humean Agents***. Oxford.

Dorst, C. (2019). “Towards a Best Predictive System Account of Laws of Nature”. ***BJPS***. 70, pp. 877–900.

Hicks, M. T. (2018). “Dynamic Humeanism”. ***BJPS*** 69.4, pp. 983–1007

Jaag, S. and C. Loew (2020). “Making Best Systems Best for Us”. ***Synthese*** 197.3, pp. 2525–2550.

Lange, M. (2009). ***Laws and Lawmakers***. Oxford.

Ward, B. (2002). “Humeanism without Humean Supervenience: A Projectivist Account of Laws and Possibilities”. ***Philosophical Studies*** 107.3, pp. 191–218.

Transparency and Trust in Science

Stephan Guttinger

University of Exeter, Exeter, United Kingdom

Transparency in science is often seen as an unalloyed good: it supports the re-use of findings and increases trust in the work of scientists more generally. Transparency is treated as a core value of Open Science, and a lack of transparency is thought to be one of the drivers of the replication crisis in the experimental sciences (Munafo et al. 2017).

However, philosophers such as Onora O’Neill (2002) and more recently C. Thi Nguyen (2022) argue that transparency has a dark side. What I will refer to as the “Nguyen-O’Neill thesis” maintains that transparency represents a damaging intrusion, especially when applied to expert domains. Pulling back the curtain makes people lie about their actions or true motives. In the case of experts, it makes them drop their normal ways of acting, because non-experts could not understand and assess them fully and fairly. Transparency might be needed to expose bias, incompetence, and corruption, but it can also leash and diminish expertise. Nguyen argues that there is no pain-free solution to this practical dilemma.

In my talk I propose that the dilemma can be dissolved. Key to this solution is to re-think what transparency is. I will show that the Nguyen-O’Neill thesis works with a static and atomistic view of transparency: the starting point is the assumption that the expert is naturally in a state of self-transparency. When faced with demands for (outward-facing) transparency, experts decide whether to disclose their reasons for action or whether to only use reasons that can be understood by non-experts. By doing so, they no longer follow their own “intimate reasons”, i.e., the trained sensitivities and awareness they acquired through experience. In such a framework, transparency represents an intrusion as it damages intimacy, and thus the epistemic power of true expert practice.

Using a case study from the life sciences, I argue that the static and atomistic view of transparency is fundamentally flawed. Rather than a state an isolated expert is in, transparency is an ongoing process that is co-produced in an active exchange with other humans, both experts and non-experts. It is only in such a dynamic and relational context that true transparency, and hence the full epistemic power of expert systems, can be realised. Rather than a damaging intrusion, the process of trans-parenting represents an infusion of insight that enhances expert systems and trust. This does not mean that trans-parenting does not pose challenges or suffer from limitations. I will conclude by reflecting on the scope and the potential downsides of this process-view of transparency in science.

- Elliott, K.C., 2022. A taxonomy of transparency in science. ***Canadian Journal of Philosophy***, 52(3), pp.342-355.
- Munafò, M.R., Nosek, B.A., Bishop, D.V., Button, K.S., Chambers, C.D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.J., Ware, J.J. and Ioannidis, J., 2017. A manifesto for reproducible science. ***Nature human behaviour***, 1(1), pp.1-9.
- Nguyen, C.T., 2022. Transparency is surveillance. ***Philosophy and Phenomenological Research***, 105(2), pp.331-361.
- O'Neill, O. 2002. A question of trust: The BBC Reith lectures 2002. Cambridge: Cambridge University Press.

CAUSAL PERSISTENCE AND LONG-RUN EFFECTS

Ina Jäntgen

University of Cambridge, Cambridge, United Kingdom

In recent years, philosophers of science have drawn attention to various properties causal relationships can exhibit, e.g., stability and specificity (Woodward 2010). These properties matter for understanding how scientists choose amongst all causes of a phenomenon those suitable for explanation or manipulation of effects.

This paper introduces a neglected but important set of properties causal relationships can exhibit to varying degrees, all connected to the persistence of causal effects:

1. ***Causing a persistent effect:*** How persistent the effect of a cause is (e.g., how persistent the cholesterol level caused by a drug is)
2. ***Causing the persistence of an effect:*** How much difference a cause makes to the persistence of its effect (e.g., how much difference the drug makes to the persistence of the cholesterol level)
3. ***Persistently causing:*** How persistent the influence of a cause over its effect is (e.g., how persistent the influence of the drug over the cholesterol level is)

I defend the importance of these properties in two steps: First, each of them tracks a different kind of control agents can exploit to bring about desired effects. Identifying these properties thus matters for understanding which causes have long-run effects particularly useful for manipulative purposes. Second, common methods used to study long-run effects do not provide sufficient evidence to attribute persistence properties to these effects. The upshot is that researchers studying long-run effects should pay more attention to providing evidence for their persistence. More precisely, I will proceed as follows:

To start, I introduce the three properties connected to the persistence of causal effects, relying on examples from biomedical research and economics (e.g., Acemoglu, Johnson, and Robinson 2001). Specifying these properties will not require adopting any particular theory of causation. However, I then analyze these persistence properties using an interventionist account of causation, following much of the literature on causal properties. This analysis allows us to see that the persistence properties are not reducible to other debated properties, most importantly not to the irreversibility and the speed of causal effects (Ross 2018; Ross and Woodward 2022).

Next, I discuss how each of the persistence properties tracks a different kind of control agents can exploit to bring about desired long-run effects. Therefore, these properties matter for selecting causes with particularly useful long-run effects, an issue of particular concern in the biomedical and social sciences.

Finally, I discuss two methods commonly used to study long-run effects: randomised controlled trials and quantitative analysis of longitudinal data. Even under ideal circumstances, these methods do not warrant an inference to the considered causal relationship being persistent in either of the specified senses. Identifying these methods as unsuitable to study causal persistence sets the ground for more discussion on which methods are suitable for this purpose. I propose some avenues to tackle this task.

Overall, this paper furthers our understanding of how to select those causes which have long-run effects particularly useful for manipulative purposes – a problem important for researchers and philosophers of science aiming to improve the study of long-run effects.

Selected references

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. “The Colonial Origins of Comparative Development: An Empirical Investigation.” American Economic Review 91 (5): 1369–1401.

Ross, Lauren N. 2018. “Causal Selection and the Pathway Concept.” Philosophy of Science 85 (4): 551–72.

Ross, Lauren N., and James F. Woodward. 2022. “Irreversible (One-Hit) and Reversible (Sustaining) Causation.” Philosophy of Science, June, 1–10.

Woodward, James. 2010. “Causation in Biology: Stability, Specificity, and the Choice of Levels of Explanation.” Biology & Philosophy 25 (3): 287–318.

BEAUTY IN EXPERIMENT: BEYOND THE CASE STUDY

Milena Ivanova

University of Cambridge, Cambridge, United Kingdom

A growing literature in philosophy of science focuses on the role of aesthetics in scientific practice and the experiment has very recently been recognized for its aesthetic value. However, the literature on aesthetics in experimentation grows out largely from case studies from the history of science. While such historical analyses highlight important ways in which aesthetic factors feature in scientific practice in the past, and the role they play in scientific reasoning, the question remains open as to whether contemporary scientists experience aesthetics in their experimental work in the same ways, given the diversity of research methods and experimental practices used today.

To address this question, we employ empirical methods to study the aesthetic attitudes scientists hold in the lab. Our approach follows a recent trend in philosophy of science to extend philosophical analysis and argumentation beyond case studies and apply experimental methods. Our work extends this approach by using interview methodology to gain insights about the role of aesthetics in scientific experimentation.

In the paper we analyse the aesthetic experiences of contemporary scientists – physicists and biologists – in experimental practice, drawing from in-depth interviews with over 200 scientists we conducted in four countries. We find that aesthetics feature in contemporary physicists' and biologists' experimental practice in much the same way as in historical cases that have recently become the focus of philosophical analysis on the beauty of experiments. Our analysis shows that in line with findings based on historical case studies, contemporary scientists experience aesthetics in engaging with the phenomena they study in the lab, designing their experiments, the data they collect and interpret, and in the significance they attribute to beauty in their field. Unlike what the historical cases studies suggest, however, we found that contemporary scientists often find performing experiments to be less aesthetically pleasing, which itself underscores ways they do experience beauty in their work.

Using the conclusions from our study, we draw some implications for current research on the aesthetics of science. So far in the debate much attention has been paid to whether aesthetic judgments in science are stable or changing. The answer to this question has serious implications on how we see the role of such values in science, whether they contribute positively to epistemic goals or undermine them. Our study gives new grounds to question claims about 'aesthetic revolutions' since those are usually defended using theoretical transitions as examples rather than experimental practice. We argue that we see surprising continuity in the appreciation of aesthetic values in different experimental traditions. Importantly, this study helps us understand the role of such values in the daily life of scientists, significantly enriching the literature by leading to a number of previously unexplored questions.

A tension between pragmatic Humeanism and realist metaphysics

Callum Duguid

University of Leeds, Leeds, United Kingdom

Humeanism is currently one of the leading approaches to natural laws. According to this view, the laws are especially efficient summaries of the events; their role is to ***describe*** nature rather than ***govern*** it. Recent defenders of this view have advocated for a pragmatic justification of the standards which pick out the laws (Hicks 2018; Dorst 2019; Jaag and Loew 2020). Dorst (forthcoming) has pushed this position further, arguing that pragmatic Humeanism contains within it a dissolution of the measurement problem of quantum mechanics.

The central idea is that the measurement problem is motivated by an anti-Humean assumption: the contradiction between a system's Schrödinger dynamics and its behaviour under observation is only apparent. It is based on treating the dynamics as rules for nature to follow rather than as guides for creatures like us. Humeans have independent reason to reject that conception of laws and so are well-motivated to see no contradiction between these two kinds of dynamics. The result is that textbook quantum mechanics emerges as a contender for the best systematisation of quantum systems, with the Schrödinger equation and the collapse postulate serving as its laws.

In this paper, I suggest that Dorst's proposal will lead to trouble for pragmatic Humeanism. The pragmatic turn threatens an epistemological link between the laws and the mosaic, which contemporary Humeanism relies on. A representationalist approach to our scientific theories justifies our use of them as guides to metaphysical ontology. However, a pragmatic view of laws arises from treating scientific theories as being primarily focused on non-representationalist objectives, such as prediction. Without the epistemological link provided by representationalism, the pragmatic Humean has no naturalistic way of filling in their ontology.

The pragmatic Humean is not alone in arriving at this position. Healey's (2017) pragmatic approach to quantum physics is also non-representationalist and rejects attempts to discover a distinctively quantum ontology. Despite similarities between the views, the comparison does not offer a way out of trouble for the pragmatic Humean. First, Healey sees reason to treat quantum physics differently to other (representationalist) physical theories, whereas the pragmatic Humean cannot take guidance on ontology from any scientific theory. Second, Healey's company is cold comfort to most Humeans, given that – from his perspective – there is little reason to engage in the heavily metaphysical project of providing reductive analyses.

The result is a view whose pragmatist approach to science is at odds with its realist approach to metaphysics. This is an instability which will be settled by either pulling back on the pragmatic influences and tackling the measurement problem head-on, or by embracing those influences and jettisoning as baggage the Humean metaphysical project.

Dorst, Chris. forthcoming. “There Is No Measurement Problem for Humeans.” ***Noûs***.

Dorst, Chris. 2019. “Towards a Best Predictive System Account of Laws of Nature.” ***The British Journal for the Philosophy of Science*** 70 (3): 877–900.

Healey, Richard. 2017. ***The Quantum Revolution in Philosophy***. New York: Oxford University Press.

Hicks, Michael Townsen. 2018. “Dynamic Humeanism.” ***The British Journal for the Philosophy of Science*** 69 (4): 983–1007.

Jaag, Siegfried, and Christian Loew. 2020. “Making Best Systems Best for Us.” ***Synthese*** 197 (6): 2525–50.

WHAT IS THE NATURE AND FUNCTION OF PRAGMATIC UNDERSTANDING?

Oscar Westerblad

University of Cambridge, Cambridge, United Kingdom

Discussions of scientific understanding centre around either explanatory understanding (EU) — understanding why — or objectual understanding (OU), understanding of a subject-matter. This focus neglects a third kind of understanding that is also recognised in the literature, but which remains underdeveloped: pragmatic understanding (PU). In this paper, I provide an account of the nature and function of PU in the context of scientific inquiry. The overall aim is to provide a fuller picture of what scientific understanding is by also providing an account of PU that can rival accounts of EU and OU.

The paper will proceed in three parts. Briefly, to set the scene, I provide an overview of undercurrents in the literature on scientific understanding that point towards PU. Many philosophers hint at PU without giving a full characterisation of it, like Peter Lipton suggesting that understanding might flow ‘abilities or skills rather than from explanations’, encouraging philosophers to take up the task ‘to point out the extent to which scientific understanding is understanding how rather than understanding why’ (2009, 54). Henk de Regt (2017, 2020) has done much to expand on PU, but unduly restricts his account to understanding of theories (UT). With this background, it becomes clear what the task of providing a full account of PU becomes: PU is a matter of understanding how, which depends on the abilities and skills of scientists. The rest of the paper explicates the nature and function of PU against the backdrop of a historical case.

In the second part of the paper, I discuss a case from the history of electromagnetism, drawing on the work of Friedrich Steinle (2016) in showing how André-Marie Ampère got a practical grip on perplexing electromagnetic phenomena through exploratory experimentation. Ampère developed particular abilities and skills in dealing with electromagnetic phenomena, phenomena that could not be explained through the theoretical frameworks available at the time. Ampère provided the foundations for methods that stabilised the electromagnetic phenomena and concepts that captured the stable regularities thereby produced, methods and concepts that were presupposed in future explanatory and theoretical work. Here, we have the beginnings of an account of PU in terms of manipulative abilities, which function to bring about methods and concepts that make sense of phenomena.

Finally, I articulate my account of PU against the backdrop of the historical case. The nature of PU is explicated in terms of the ability to manipulate a target system, where these abilities are learned and honed, so that they constitute skills that embody methods (cf. Bengson 2020). By interacting methodically with nature, scientists develop and subsequently apply concepts that make phenomena intelligible and better understood. PU thus has a distinctive function in the sciences by grounding scientists' methodical interactions with and conceptual articulations of nature. PU helps us capture the epistemic achievements and understanding of Ampère and others like him, which then provides a fuller understanding of scientific understanding, filling in details left out by EU, OU, and those who have hinted at the relevance of PU.

Indicative bibliography

Bengson, John. (2020). 'Practical Understanding: Skill as Grasp of Method', in Demmerling, Christoph; Schröder, Dirk (eds.) Concepts in Thought, Action, and Emotion, Routledge Press: 215-235.

De Regt, Henk. (2017). Understanding Scientific Understanding, Oxford: Oxford University Press.

De Regt, Henk. (2020). 'Understanding, Values, and the Aims of Science', Philosophy of Science, 87: 921–932.

Lipton, Peter. (2009). 'Understanding Without Explanation', in Henk de Regt, Sabina Leonelli, and Kai Eigner, Scientific Understanding: Philosophical Perspectives, Pittsburgh: Pittsburgh University Press: 43-63.

Steinle, Friedrich. (2016). Exploratory Experimentation: Ampère, Faraday, and the Origins of Electrodynamics, Pittsburgh: University of Pittsburgh Press.

MUST THE CAUSAL STRUCTURE OF THE WORLD BE UNIQUELY IDENTIFIABLE?

Veli-Pekka Parkkinen

University of Bergen, Bergen, Norway

In the most general terms, a causal structure comprises a set of causal relata --- the structure's component parts --- and a causal ordering over them. Given a causal ordering, each part can be classified as either endogenous or exogenous: either an effect of other parts of the structure, or uncaused by other parts. It is impossible to study the causal structure of the whole world at once. Hence, a causal structure as a target of empirical analysis is implicitly defined relative to a causal context, which itself is just a larger causal structure omitted from the description of the analyzed structure.

A causal structure is qualitatively uniquely identifiable if, and only if, it is possible to generate data about the joint behavior of its parts such that those data determine a unique causal ordering over the parts. Assumptions about qualitative causal ordering are in turn required to estimate the size of particular causal effects between parts of the structure. Often, data generated by observation underdetermines causal ordering, and the analyzed structure is not uniquely identified without causal assumptions. When possible, identification may be aided by experiment: an experimental manipulation is a known exogenous cause that is leveraged to resolve observationally ambiguous structure.

I argue that various philosophical theories of causation, as well as methods for causal discovery and estimation of quantitative causal effects, involve assumptions that are equivalent to assuming that every causal structure is in principle uniquely identifiable without any causal assumptions or knowledge. The basis of the argument is the following "unique identifiability thesis": If every endogenous part of a causal structure has at least one exogenous cause that is independent of other exogenous causes, or the structure can be embedded into a larger structure including such exogenous causes, then the structure is uniquely identifiable without causal assumptions. This thesis has been demonstrated by Spirtes et al. (Spirtes et al. 2000, pp. 64-65): when the endogenous parts of a structure have mutually independent exogenous causes, the resulting structure of associations determines unique qualitative causal ordering.

The rest of my argument is based on an observation that various theories of causation and methods for causal inference and discovery assume the antecedent of the aforementioned thesis. Structural equation modeling, and more broadly, regression-analytic methods assume independent (causal) sources of error for the dependent variable(s). Papineau (2022) defines causation as a dependence tracked by structural equations with independent errors, excluding everything but uniquely identifiable structures from the extension of the concept of causal structure. Spirtes et al. (2000) assume unique identifiability whenever they assume pseudo-indeterminism. Woodward (2003) assumes unique identifiability by assuming interventions that eliminate confounding for every cause-effect pair. Possibly even deterministic theories of causation (e.g. Baumgartner & Falk, 2019) make an analogous assumption about exogenous independence, with the same consequence.

To assume that every causal structure is uniquely identifiable puts very particular constraints on what the causal structure of the whole world must be like. That is, it is not a realistic assumption. Rather, I contend that the ubiquity of such an assumption perhaps indicates that causal reasoning primarily applies to, or is particularly useful for analyzing locally bounded systems that can be decomposed to independently controllable parts, and is less useful elsewhere.

References:

Baumgartner, M., & Falk, C. (2019). Boolean difference-making: a modern regularity theory of causation. The British Journal for the Philosophy of Science.

Papineau, D. The Statistical Nature of Causation (2022), The Monist, 105(2), 247–275.

Spirtes, P., C. N. Glymour, R. Scheines, and D. Heckerman (2000). Causation, prediction, and search. Cambridge, MA: MIT press.

Woodward, J. Making Things Happen (2003). Oxford: Oxford University Press.

INTEGRATED HISTORY, PHILOSOPHY, AND SOCIAL STUDIES OF SCIENCE

AN APOLOGY FOR ANALYTIC PHILOSOPHY: THE LEFT VIENNA CIRCLE IN POSTWAR AMERICAN ACADEMIA

Noah Friedman-Biglin

San José State University, San José, USA

In its heyday of the late 1920s and early 1930s, Logical Positivism (as it was then known) was a movement which brought a program of philosophical research together with an active mission of societal reform, especially in the hands of the Vienna Circle. Their ambitious program of public outreach, via lectures in the Verin Earnst Mach (The Earnst Mach Society) and engagement with other sympathetic groups like the Dessau Bauhas and Berlin Circle, was matched by the strident tones of their writing and vigor of the debates in their meetings. But, as fascism began to rise in Europe, the environment in Vienna became hostile to their movement; it was not lost on the new fascist governments that Logical Positivism was allied with left-wing political projects, and with a general anti-fascist sentiment. By 1936 the situation for the members of the Vienna Circle became untenable: Moritz Schlick was murdered on the steps of the University of Vienna, and the rest of the Circle began to make plans to flee.

Several members of the Circle managed to immigrate to the United States where they had an immense impact on the shape of American academia. But, even though some of the members of the Circle continued their work in their adopted country, the movement never again regained the crusading spirit. To paraphrase Reisch, they transformed their fiery rhetoric "to the icy slopes of logic", and never again had an active association with politics. It further transformed, first into Logical Empiricism, and finally into the more generic Analytic Philosophy. It became a methodology safely quarantined in academic philosophy, and far from any influence on the society of post-war America.

What explains this transition is the subject of this paper. I will argue that, contrary to the history outlined above, at least some of the members of the Circle never lost their ambition that the philosophical movement they started would have far ranging societal consequences. Or, in other words, that it would have political outcomes. I will focus on the case of Carnap, and in particular, on his pluralistic approach to logic. I will show that the tack he takes in debates about logic, and later about ontology, is driven by what he called "The Scientific World Conception", a platform of ideals he laid out in the Circle's manifesto, and that he still held in the post- war era. Finally, I will argue that the reason that the political ambitions of Carnap's work are largely unknown in analytic philosophy today stem from the political environment in academia in the 1950s, and the hostility to left-wing projects at the time. According to my view, a conscious effort was made by him to downplay this aspect of the Circle's project to avoid scrutiny by the American government, and as a result of his experience with fascism in Europe.

1. Carnap, R. (1929). “Wissenschaftliche Weltauffassung: Der Weiner Kreis (The Scientific World-conception: The Vienna Circle)”. In Cohen, R. and Neurath, M. (eds.) Empiricism and Sociology, pp 299 – 318. Dordrecht: D. Reidel Publishing Company.

2. Carnap, R. (1937). The Logical Syntax of Language. London: Kegan Paul, Trubner & Co.

3. Reisch, George A. (2005). How the Cold War Transformed the Philosophy of Science: To the Icy Slopes of Logic. Cambridge: Cambridge University Press.

4. Reisch, George A. (2019). The Politics of Paradigms: Thomas S. Kuhn, James B. Conant, and the Cold War Struggle for Men's Minds. Albany, NY: SUNY Press.

5. Uebel, T. (2005). “Political Philosophy of Science in Logical Empiricism: the Left Vienna Circle”. Studies in History and Philosophy of Science, Vol. 36: 754 – 773.

Negative Analogies and Representation in Medical Practice: A Case from Clinical Orthopaedics

Helene Scott-Fordsmand[1], Mauricio Suárez[2]

[1]University of Cambridge/Clare Hall, Cambridge, United Kingdom. [2]Clare Hall, Cambridge, United Kingdom

We argue for a role for analogical reasoning in medical practice, in particular in clinical orthopaedics, and show that exploring this role can contribute to the philosophical understanding of scientific representation. Our argument proceeds by means of a case study in the use of classificatory representations in treatment of shoulder fractures. A key device in diagnostics in this area is the Neer classification schema, dating from 1970 (Neer 1970). This schema looks at face value like an isomorphic representation of natural classes of bone fracture, with each of sixteen distinct graphical depictions holding a relation of positive similarity in shape and displacement to their appropriate targets in the patient population. However, in practice, employing the Neer classification is far more complex than merely matching patient cases to the fracture classes. Instead, the system works heuristically, in terms of guiding attention and action – even in cases where there is no obvious conclusion as to which class the patient case belongs to. Drawing from observations of clinical practice and sources on the historical development of the Neer classification schema we show, first, that it functions not as a pictorial depiction, but as a scientific model. And that taking it as such makes sense of clinical examples as well as statements made by Neer on the nature of his classification system. Rather than a depiction of natural classes, the Neer system can thus be seen as a template for analogical reasoning. Second, we argue that the intuitive appeal to understand this classificatory system – and others like it – in terms of similarity between class and case is particularly misleading in that it overlooks the importance of relations across entire classificatory systems, as well as patterns of reasoning through negative analogies. Drawing on Mary Hesse’s tripartite distinction between positive, neutral, and negative analogies, we show that, contrary to received wisdom, negative analogies play as important a role, if not the decisive role, in how the model supports reasoning and decision making. Our analysis is of a piece with other recent studies that emphasise the indispensable role of ‘inverse negative analogies’ (Pero and Suárez, 2016) and ‘dissimilarities’ (Boesch, 2021) in scientific modelling practice. These analyses display the subtle ways in which the different elements in a model interact holistically in inferential practice (Fang, 2019, Suárez, 2010), which is rendered opaque by analyses in terms of similarity, or isomorphism.

References:

Bartha, Paul. 2021. “Analogy and Analogical Reasoning”, ***Stanford Encyclopaedia of Philosophy.*** https://plato.stanford.edu/entries/reasoning-analogy/

Black, Max. 1962. ***Models and Metaphors.*** Ithaca: Cornell University Press.

Brorson, Stig and L. Henrik Frich and A. Hrobjartsson. 2009. “The Neer classification for fractures of the proximal humerus: a narrative review”, ***Minerva Ortopedica e Traumatologica,*** 60(5), pp. 447-60

Boesch, Brandon. 2021. “Scientific Representation and Dissimilarity”, ***Synthese***, 198, pp. 5495-5513.

Fang, Wei. 2017. “Holistic modeling: an objection to Weisberg’s weighted feature-matching account”. ***Synthese***, 194, pp. 1743–1764.

Hesse, Mary. 1966. ***Models and Analogies in Science.*** Notre Dame: Notre Dame University Press.

Neer, Charles S. 1970. “Displaced Proximal Humeral Fractures”, ***Journal of Bone and Joint Surgery,*** 52A, pp. 1077–1089.

Neer, Charles S. 2002. “Four-segment classification of proximal humeral fractures: Purpose and reliable use”, ***Journal of Shoulder Elbow Surgery,*** July/August, pp. 389-400

Pero, Francesca, and Mauricio Suárez. 2016. “Varieties of Misrepresentation and Homomorphism”. ***European Journal for Philosophy of Science***, 6(1), pp. 71-90.

Suárez, Mauricio. 2010. “Scientific Representation”. ***Philosophy Compass***, 5 (1), pp. 91-101.

ON THE VALUE AND CHALLENGES OF PLURALISM IN SCIENCE: FEYERABEND AND BOHM IN BRISTOL

Marij van Strien

Institut Wiener Kreis, Vienna, Austria

Quantum mechanics is often regarded as an important case for scientific pluralism, as the theory allows for a plurality of interpretations. This talk shows that scientific pluralism is also historically connected to quantum mechanics: in particular, Paul Feyerabend developed his pluralism through his engagement with quantum mechanics and in conversation with the physicist David Bohm, whom he met in Bristol in 1957. However, the pluralism which Feyerabend and Bohm argued for turned out to be hard to realize in practice.

In 1952, Bohm published an alternative interpretation of quantum mechanics, demonstrating the possibility of non-standard interpretations. However, Bohm himself regarded this interpretation merely as a starting point, and argued that what was needed was the development of new concepts, which could form the basis for a genuinely new theory yielding new predictions. In this context, Bohm developed general arguments for pluralism in science: to avoid being trapped within a conceptual scheme, scientists should actively try to develop alternatives to current theories.

For Feyerabend, the foundations of quantum mechanics were a main research focus throughout the 1950s and 1960s (Van Strien 2020, Kuby 2021). In 1957, Feyerabend and Bohm both participated in a conference in Bristol on the foundations of quantum mechanics (Kožnjak 2018). Later that year, they became colleagues in Bristol, where they regularly discussed physics and philosophy. The fact that Bohm had a large influence on the development of Feyerabend’s pluralistic philosophy of science can be seen from the fact that Feyerabend attributed one of his main arguments for pluralism to Bohm. He saw pluralism as particularly urgent in quantum physics, where, in his perception, Bohm’s interpretation was dogmatically rejected by the scientific community.

Thus, Feyerabend’s scientific pluralism emerged out of an engagement with concrete issues in science. However, as Feyerabend’s understanding of the complexities of quantum physics and its historical development grew, his criticism of the standard interpretation of quantum mechanics gradually became weaker. He came to appreciate Niels Bohr as a creative scientist, and even as an exemplary anarchist: Feyerabend repeatedly stated that it was his engagement with Bohr’s work on quantum physics which led him to an anarchist position in philosophy of science (Van Strien 2020). Although in Against Method, Feyerabend presented pluralism as a virtue, he no longer thought that it should be imposed on science as a methodological requirement.

Meanwhile, Bohm’s attempts to develop a new conceptual framework for quantum physics remained largely unsuccesful, and from the late 1970s, he returned to his original interpretation from 1952. This interpretation has become increasingly popular, but it is not the genuinely new theory which Bohm envisioned: it largely uses classical concepts and has not yielded new predictions. Despite the plurality of interpretations of quantum mechanics which one can find nowadays, it is hard to find one which presents a new theoretical framework in the way Bohm and Feyerabend envisioned. It thus seems that while scientific pluralism has been developed within specific scientific contexts, scientific practice has also set limits to pluralism.

- Kožnjak, B. 2018. "The missing history of Bohm's hidden variables theory: The Ninth Symposium of the Colston research society, Bristol, 1957." ***Studies in History and Philosophy of Modern Physics,*** 62: 85–97
- Kuby, D. 2021. "Feyerabend's Reevaluation of Scientific Practice: Quantum Mechanics, Realism and Niels Bohr." In ***Interpreting Feyerabend: Critical Essays,*** ed. K. Bschir & J. Shaw. Cambridge University Press.
- Van Strien, M. 2020. "Pluralism and Anarchism in Quantum Physics: Paul Feyerabend's Writings on Quantum Physics in Relation to his General Philosophy of Science." ***Studies in History and Philosophy of Science*** 80: 72-81.

COSTS OF PLURALISM

Teemu Lari, Uskali Mäki

University of Helsinki, Helsinki, Finland

Plurality and diversity in science have been associated with significant epistemic benefits. Knowledge-extending benefits may result when a plurality of models, theories, or methods can jointly serve a wider range of epistemic purposes than any of them alone (Giere, 2006; Mitchell, 2003). Knowledge-improving benefits may result from interaction between researchers who differ in their knowledge, perspectives, background commitments, skills, and other factors, when criticism and debate help remove biases and mistakes (Longino, 2002). Some authors stress that researchers should even actively try to increase the degree of plurality and diversity in science (Chang, 2012).

However, an increased degree of plurality and diversity does not only have benefits but also drawbacks or costs, which have only been explored in a cursory way. Any judgment on whether some disciplines should have more plurality or diversity of some kind – whatever the means to effect such a change – one should also consider the costs. In this paper, we develop a more comprehensive and detailed understanding of the costs of pluralism, proceeding in the spirit of the economics of scientific knowledge (Zamora Bonilla, 2012) and consulting the fields of transaction cost economics (Williamson, 2008) and organizational sociology of science (Whitley, 2000).

We begin by cataloguing the epistemic benefits that various kinds of plurality and diversity may have according to the philosophical literature. Each presumed benefit may increase the need for scientific labour in one form or another. Costs result from 1) the labour needed to bring the potentially beneficial plurality into existence in the first place. For example, it is more economical to use one model template to represent a set of target phenomena, compared to developing a range of altogether unrelated, and thus more diverse models for the same purpose. Costs also result from 2) the labour needed to process the plurality and diversity into epistemic benefits. For example, the supposed benefits of criticism across theoretical or disciplinary perspectives only materialize when scientists dedicate enough time and effort to making the exchanges informed and constructive. Finally, costs result from 3) the labour needed to counter adverse effects of plurality and diversity. For example, some forms of plurality might weaken the scientific consensus, and if the perception of consensus by the public supports trust in science, then increased plurality means scientists must work harder to maintain public trust in science.

Transaction cost economics explains existing institutions as cost-minimizing arrangements, given any level of benefits. In the case of plurality and diversity in science, cost-benefit comparisons are difficult since the costs and benefits do not seem commensurable. Expected benefits are epistemic, indirect, and more uncertain, while the costs are non-epistemic, more direct, and less uncertain. We will examine this observation and its implications. In particular, we ask whether the high intellectual transaction costs that accompany high degrees of plurality and diversity could explain differences in the actual degree of them in various disciplines, given differences in the institutional aspects of the disciplines.

Chang, H. (2012). Is water H_2O? Evidence, Realism and Pluralism. Springer.

Giere, R. (2006). Scientific Perspectivism. University of Chicago Press.

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

Mitchell, S. (2003). Biological Complexity and Integrative Pluralism. Cambridge University Press.

Whitley, R. (2000). The intellectual and social organization of the sciences (2nd ed.). Oxford University Press.

Williamson, O. (2008). Transaction Cost Economics. In C. Ménard & M. Shirley (Eds.), Handbook of New Institutional Economics (pp. 41–65). Springer.

Zamora Bonilla, J. (2012). The Economics of Scientific Knowledge. In U. Mäki (Ed.), Philosophy of Economics (pp. 823–862). Elsevier.

Philosophy of the Cognitive Sciences

Neural Representations Unobserved

Marco Facchin

IUSS Pavia, Pavia, Italy

From single cell recordings to multivariate pattern analysis, the experimental techniques of cognitive neuroscience are commonly said to reveal the neural basis of various representational phenomena. By relying on such techniques, we discovered that individual neurons represent their "preferred" stimuli, that cortical maps represent what they are homomorphic to, and that the activation spaces of various neural structures are dynamic models of various external states of affairs. Cognitive neuroscience, then, shows us representations made flesh - the neural vehicles underpinning our cognitive processing and representational capacities [e.g. 1].

Here, I wish to contest this familiar picture. I want to argue that the neural structures cognitive neuroscience shows us are not, to the best of our scientific knowledge, representational vehicles.

To claim so, I will first sketch a functional characterization of representational vehicles that is standardly accepted in philosophy of cognitive neuroscience. On this view, representational vehicles are, first and foremost, action-guiding structures that structurally resemble the targets they represent. This means, amongst other things, that the representational vehicles must be (roughly) homomorphic to their targets, and that the degree to which they are homomorphic to the target must correlate with an agent's chances of behavioral success: the better the homomorphism, the higher the odds of success [e.g. 1,2].

Equipped with this account of representational vehicles, I will argue that individual neuronal responses, neural maps and activation spaces do not qualify as representational vehicles. My argument will be simple. I will show that, whilst neural maps are homomorphic to what they represent, this homomorphism does not play the required action-guiding role. This is because it does not play a role in determining the input/output profile of neural maps. As such, it does not influence the agent's behavior as the characterization of representational vehicles requires. Indeed, contrary to what the action-guidance predicts, it has been observed that learning processes that increase an agent's odds of success worsen the homomorphism holding between the neural map and its target [cf. 3].

I will also argue that, whilst it is possible to identify an homomorphism holding between individual neuronal responses (and activation spaces) and their targets [e.g. 4], this homomorphism is not adequate to satisfy the relevant functional profile. This is because it holds amongst several different neuronal responses - each of which is taken to be a representational vehicle in its own right - and their targets [cf. 5]. So, whilst there is a relevant homomorphism, and that homomorphism plays the relevant action guiding role, it does not hold between individual vehicles and individual targets as required by the relevant account of representational vehicles.

Having shown that bona fide neural representational vehicles fail to actually qualify as such according to a standard account of representational vehicles, I will conclude my talk exploring the implication of my claim when it comes to mechanistic explanations in cognitive neuroscience [cf. 1].

[1] Piccinini, G. (2020). Neurocognitive Mechanisms. New York: Oxford University Press

[2] Gładziejewski, P. (2016). Predictive coding and representationalism, Synthese, 193(2), 559-582.

[3] Martin, J. H., Engber, D., & Meng, Z. (2005). Effect of forelimb use on postnatal development of the forelimb motor representation in primary motor cortex of the cat. Journal of neurophysiology, 93(5), 2822-2831.

[4] Nirshberg, G., & Shapiro, L. (2021). Structural and indicator representations: a difference in degree, not kind. Synthese, 198(8), 7647-7664.

[5] Connolly, A. C., et al. (2012). The representation of biological classes in the human brain. Journal of Neuroscience, 32(8), 2608-2618.

NON-SYMBOLIC FEW-SHOT LEARNING

Nina Poth

Humboldt University, Berlin, Germany. DFG Excellence Cluster "Science of Intelligence", Technical University Berlin, Berlin, Germany

A central feature often associated with general intelligence in contemporary cognitive science is few-shot learning: the ability to find the appropriate solution to a learning task with one or only a few trials. Experiments in animal science and developmental psychology explain this ability as a reflection of imagination, systematic thought, and creativity in New Caledonian crows and as a key aspect of concept learning in children; it also serves as a benchmark to build human-like AI (Halina 2021). These interdisciplinary perspectives agree that some form of mental-model building is involved, but the precise mechanisms underlying few-shot learning remain insufficiently understood.

This talk identifies as a key source of this issue that researchers persistently rely on symbolic approaches to representation and computation without looking to available alternatives. Leading research on concept formation in Bayesian AI illustrates this. Modelers interested in few-shot learning claim that "inference over hierarchies of probabilistic generative programs offers a normative and descriptive account of children's model building" (Ullman & Tenenbaum 2020). At its heart, this framework utilizes a theory-theoretic understanding of concepts, and is taken to motivate and justify a recent revival of the language of thought hypothesis as "The Best Game in Town" (Quilty-Dunn et al. 2022). Here I challenge this narrative and argue that these recent advances in computational modeling equally well support an alternative perspective on mental-model building. Specifically, I suggest using Conceptual Spaces

(Gärdenfors 2000) as a framework for representing perceptual knowledge in the form of information about similarities among sensorimotor interactions with the world; the systematic re-combination of this information in reasoning follows principles of geometric organization, as opposed to an innate syntax. This application suggests non-symbolic conceptual representations as underlying few-shot learning. Thus, as a plausible alternative, it calls into question dominating language-of-thought-like interpretations of the phenomenon.

This result is significant for the prospects of drawing from cognitive science to understand intelligent systems. The reviving focus on symbols invites theoretical ignorance towards familiar problems with symbol-grounding and negligence towards embodiment, action-oriented processing and intentionality. This makes it unsurprising that these ingredients remain lacking in pioneering AI models (Strasser & Wilby 2023). While Quilty-Dunn et al. (2022) attempt to "refresh the dialectic" under a "pluralist perspective" on mental formats and processes, relying on a language of thought as focal framework risks returning to classical dichotomies between abstract reasoning on the one hand, and perceiving and acting in the world on the other. A similarity-spaces view promises theoretical progress away from a `sandwich model', and better captures efficient learning and reasoning with perceptual concepts and skill as a form of motor intelligence (Fridland 2016). It thereby allows reinterpreting recent work in probabilistic modeling in light of embodied and action-oriented perspectives to inform a more ecumenical understanding of few-shot learning.

References

Fridland, E. (2017). Skill and motor control: Intelligence all the way down. ***Philosophical studies, 174***, 1539-1560.

Gärdenfors, P. (2000). ***Conceptual spaces: The geometry of thought***. MIT press.

Halina, M. (2021). Insightful artificial intelligence. ***Mind & Language, 36***(2), 315-329.

Strasser, A., & Wilby, M. (2023). ***The AI-Stance: Crossing the Terra Incognita of Human-Machine Interactions?*** In Raul Hakli, Pekka Mäkelä & Johanna Seibt (eds.), Social Robots in Social Institutions. Amsterdam: IOS Press. pp. 286-295.

Quilty-Dunn, J., Porot, N., & Mandelbaum, E. (2022). The Best Game in Town: The re-emergence of the language of thought hypothesis across the cognitive sciences. ***Behavioral and Brain Sciences***, 1-55.

Ullman, T. D., & Tenenbaum, J. B. (2020). Bayesian models of conceptual development: Learning as building models of the world. ***Annual Review of Developmental Psychology, 2***, 533-558.

SKEPTICISM ABOUT COMMON KNOWLEDGE

Giorgio Sbardolini

MCMP, LMU, Munich, Germany

Something is common knowledge (henceforth, CK) in a group if and only if everyone in the group knows it, everyone in the group knows that everyone in the group knows it, and so on ad infinitum. A typical application of CK is in explaining successful collective rational action in philosophy, linguistics, economics, and computer science. Several arguments target the assumption that ordinary people may share CK, in particular Lederman (2018) and Aumann (1976). For Greco (2014), the cost of rejecting CK includes rejecting scientific theories that appeal to it. On my view, some skeptical arguments against CK are sound, but we should not reject successful scientific theories that appeal to it. My position rests on an appreciation of the explanatory role played by CK in theories of collective action, and by the ideally rational people who share it. Who are the ideally rational people?

A widely shared view is that ideally rational people are abstractions of ordinary people. Accordingly, the relationship between us and them is similar to the relationship between a table surface and the frictionless plane (Williamson, 2000): a mathematical simplification obtained by setting a negligible parameter to the limit. This view is mistaken. Ideally rational people are qualitatively different from us. Lederman's (2018) argument shows that they cannot be ignorant of each other's beliefs, and Aumann's (1976) "agreement theorem" shows that they cannot disagree with each other if they share common priors. CK is a limit that introduces discontinuities between the properties described in the model at infinity and the properties described in the model at any point below infinity. In this regard, a more appropriate analogy than the frictionless plane are infinite limits in the physics of phase transitions (Batterman, 2011).

Evidence for discontinuity comes from well-known mathematical results such as Rubinstein's (1992) "electronic mail game". In this setup, two partners profit from coordinating on different actions A and B, depending on whether ***p*** or ***not-p***. If they have CK that ***not-p***, they both choose one action, say B. If they lack CK that ***not-p***, they both choose the other action A, no matter how many finite iterations of the knowledge operator are assumed to hold. Rubinstein's game shows that CK as an infinite limit indicates a failure of reduction (in some sense of the word) between ideally rational agents with CK and ordinary people: the finitary (almost-CK) model does not approximate the behavior of the model at infinity.

Thus, ordinary people may not have CK but ideally rational agents do. It does not follow that CK is of no scientific value. Its value does not lie in its being possibly instantiated in empirical reality nor in its mathematical simplicity. We appeal to CK in models of collective rational action because the kind of variation allowed by finitary models is irrelevant to our understanding of phenomena such as coordination, just like the finite number of particles in a pot is irrelevant to our understanding of vaporization.

R. Aumann (1976). “Agreeing to Disagree”, The Annals of Statistics, 4: 1236–1239.

R. Batterman (2011). “Emergence, singularities, and symmetry breaking”, Foundations of Physics, 41: 1031–1050.

D. Greco (2014). “Could KK be OK?”, The Journal of Philosophy, 111: 169–197.

H. Lederman (2018). “Uncommon Knowledge”, Mind, 127: 1069–1105.

T. Williamson (2000). Knowledge and Its Limits. Oxford: Oxford University Press.

CAN WE EXTRACT A THEORY OF CONTENT FROM COGNITIVE SCIENCE?

Johan Heemskerk

University of Warwick, Coventry, United Kingdom

Theories of representational content aim to discover, for any given representation, the principles which determine the distal content of that representation. Many philosophers working in the field, notably those concerned with teleosemantics, look to the non-philosophical branches of cognitive science (e.g. psychology, cognitive neuroscience) to provide such a theory. As Tyler Burge writes, this approach is based on the thought that cognitive science has discovered, "without being fully aware of its own accomplishment" (Burge, 2010), a theory of content.

In this paper I reconstruct a challenge to this approach due to Frances Egan. Egan argues that cognitive science has no implicit theory of content determination. She argues that content is not included within the 'theory proper' of cognitive science. The theory proper is what does the explanatory work, while various pragmatic exigencies account for non-explanatory elements.

Specifically, Egan’s challenge consists in her argument that content talk is a ‘gloss’ designed to aid comprehension of mathematical functions carried out by the brain. As such, content attributions in scientific theories are made on the basis of pragmatic decisions such as how to best communicate the theory to readers.

In order to be considered part of the theory proper, content must be ‘essential’ to the representational structures under consideration, and content must be determined by a ‘privileged naturalistic relation’ (Egan, 2018). I endorse Egan’s restrictions. The aim of the paper is to show how cognitive science meets them. I focus on a cognitive neuroscience study by Chang and Tsao (2017).

I argue, first, that content is ‘essential’ to the representational states and structures under consideration. By essential, Egan means that representations are individuated by their content in such a way as to ‘make a difference’ to the system itself. I attempt to show that cognitive science does in fact individuate representations by content. Indeed it must, since cognitive subsystems in which the same mathematical functions are computed serve diverse functional roles in the wider system just in virtue of a difference in content.

Second, content is determined by a privileged naturalistic relation holding between a representation and its distal content. In the case I focus on, this naturalistic relation is one of ‘encoding’. Encoding is not itself a mere gloss, but comes with hefty theoretical commitments which are central to the explanation provided. This generalises, and it emerges that cognitive scientists are implicitly invoking a theory of content determination which draws heavily on the resources of communication theory.

Accepting Egan’s limitations, I close by suggesting a general schema by which to identify whether content is used within the theory proper of cognitive science; the content must be specified by way of a technical concept, the invoked relationship between the representation and a distal item must be theoretically 'rich', and the content must be explanatory of some capacity of the system in a way which `makes a difference' to the system itself. It is the work of philosophers to construct a theory of content by drawing from various studies which conform to this schema.

Tyler Burge. ***Origins of objectivity.*** Oxford University Press, 2010

Frances Egan. 'The nature and function of content in computational models.' In The Routledge handbook of the computational mind, pages 247–258. Routledge, 2018.

Karen Neander. ***A Mark of the Mental: In Defense of Informational Teleosemantics***. MIT Press, 2017.

Le Chang and Doris Y Tsao. 'The code for facial identity in the primate brain.' Cell, 169(6):1013–1028, 2017.

Shea, Nicholas. ***Representation in cognitive science***. Oxford University Press, 2018.

Representational similarity analysis underdetermines similarity of object recognition mechanisms in deep neural networks and the brain

Bojana Grujicic

Max Planck School of Cognition, Leipzig, Germany. Humboldt-Universität zu Berlin, Berlin School of Mind and Brain, Berlin, Germany. University College London, Department of Science and Technology Studies, London, United Kingdom

Recent findings in visual neuroscience suggest that deep convolutional neural networks (DCNNs) trained in an object recognition task enable predicting neural response properties in the ventral stream in human brains to a certain extent (Khaligh-Razavi & Kriegeskorte, 2014). Regarding their object recognition task performance, DCNNs are on the human performance level. Based on these findings DCNNs are said to be the most predictively successful models of the ventral stream responsible for object recognition (Cao & Yamins, 2021a, 2021b).

Given these findings about their predictive success, do DCNNs also provide an explanation of our capacity of object recognition? This issue has been recently picked up in the philosophical discourse, with several arguments offered for the claim that DCNNs are mechanistic explanations of object recognition (Cao & Yamins, 2021a, 2021b; Buckner, 2018).

A key issue with such claims is a lack of critical engagement with the frameworks used to assess similarity of processing of DCNNs and the brain, which form the backdrop against which one may ascribe the mechanistic explanatory status to DCNNs. Besides linear mapping, another frequently utilised framework is that of representational similarity analysis (RSA), that compares the degree of similarity between representational geometries of DCNNs and the brain in the object recognition task (Khaligh-Razavi & Kriegeskorte, 2014).

I focus on RSA and ask whether it enables an abstract mechanistic mapping between DCNNs and the ventral stream representational mechanism (Bechtel, 2007) responsible for object recognition. I outline an account of mechanism sketches based on Craver & Kaplan (2020), and argue that RSA does not corroborate DCNNs as mechanism sketches. The issue that plagues the applications of RSA is the variety of similarity measures used as a part of that framework, such as correlation, cosine, Euclidean and Mahalanobis distances. Focusing in particular on correlation and Euclidean distance I show that they pick out different properties of stimuli-elicited patterns in order to quantify representational geometries. On a background of relevant neuroscientific evidence, I show that this further entails contradictory implications about the vehicles of representations according to two accounts of representational mechanisms one may want to map via RSA – one on the level of individual neurons comprising neural populations (Cao & Yamins, 2021a), and another on the level of neural manifolds (Buckner, 2018).

I argue that there is a problem of relevance of these similarity measures for the explanandum capacity of object recognition. Mechanisms are meant to produce, maintain or underly capacities to be explained

(Craver & Kaplan, 2020), and the features of stimuli-elicited patterns RSA picks out to quantify representational geometries have to be relevant for object recognition. Since there is currently no arbitration between similarity measures in the field in terms of relevance for object recognition, RSA underdetermines DCNNs as abstract mechanistic models of object recognition.

Bechtel, W. (2007). Mental mechanisms: Philosophical perspectives on cognitive neuroscience. Psychology Press.

Buckner, C. (2018). Empiricism without magic: transformational abstraction in deep convolutional neural networks. Synthese, 195(12), 5339-5372.

Cao, R., & Yamins, D. (2021a). Explanatory models in neuroscience: Part 1--taking mechanistic abstraction seriously. arXiv preprint arXiv:2104.01490.

Cao, R., & Yamins, D. (2021b). Explanatory models in neuroscience: Part 2--constraint-based intelligibility. arXiv preprint arXiv:2104.01489.

Craver, C. F., & Kaplan, D. M. (2020). Are More Details Better? On the Norms of Completeness for Mechanistic Explanations. British Journal for the Philosophy of Science, 71(1), 287-319.

Khaligh-Razavi, S.-M., & Kriegeskorte, N. (2014). Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. PLoS Computational Biology, 10(11), e1003915.

On the Benefits of 'Hand' Engineering in Neuroscience

Nedah Nemati

Columbia University, New York, USA

Over the past two decades, neuroscience has seen a shift away from methods of 'hand' engineering – or manually picking specific, relevant components required to produce an action – toward a more automated and abstract use of patterns generated from many data points, or what I call 'pattern' engineering. This shift, most notable in theoretical and computational neuroscience, is evidenced by the adoption of various methods, including the use of multivariate analyses (Cox and Savoy 2003; Norman et al. 2006; Kriegeskorte et al. 2008), the move from individual neural responses to population dynamics (Abbott and Dayan 1999; Averbeck et al. 2006), and new statistical approaches taken with deep learning methods (Kietzmann et al. 2017).

One way to settle the dispute between 'hand' and 'pattern' engineering is to empirically compare their approaches. However, there is no shared basis for such comparisons, as the aims of researchers can differ widely. For example, while hand engineering has been criticized for its poor performance, what constitutes 'performance' itself has remained elusive, indexed to the various tasks set by researchers.

Some scholars, identifying shifts in the epistemic benefit of these newer methods, have also noted how scientists have often defined performance as prediction (Chirimuuta 2021).

Nonetheless, behavioral work in neuroscience aims to do more than simply build models that predict behavior. In addition to being a resource for explaining human behavior, behavioral neurobiologists attempt to integrate results within other areas of scientific practice. For example, while the tradeoff between prediction and intelligibility might present one kind of challenge, other explanatory aims in behavioral neuroscience – such as deriving explanations that scale across species – bring their own sets of issues. Moreover, the move away from hand engineering has more recently permeated biological queries, with neurobiologists turning to such ‘pattern’ approaches to guide large-scale, transcriptomic data sets. This work explores what else is lost for explanations of behavior as neuroscientists rely on pattern matching, and what consequences such losses may have in other areas of neuroscience that explicitly draw on biological properties to characterize behavior.

The first part of the paper discusses the epistemic issues raised by the shift from ‘hand’ to ‘pattern’ engineering in light of both computational tools in neuroscience and specific examples in behavioral neurobiology. I next engage the intersections between these areas of neuroscience by discussing how computational models can be both biologically inspired and subsequently imported back into biology. Contrary to conventional belief, my analysis shows neglected benefits to 'hand' engineering, even in the face of its poorer predictive capabilities. I argue that this neglect is only apparent through conceptual as opposed to empirical methods, given that the latter are only capable of understanding performance as a proxy for explanation rather than explanation within broader contexts of behavioral understanding.

Chirimuuta, M. Prediction versus understanding in computationally enhanced neuroscience. Synthese 199, 767–790 (2021).

Cox DD, Savoy RL. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage. 2003 Jun;19(2 Pt 1):261-70.

Kietzmann, Tim C., Patrick McClure, and Nikolaus Kriegeskorte. "Deep neural networks in computational neuroscience." BioRxiv (2017): 133504

Kriegeskorte N, Mur M, Bandettini P. Representational similarity analysis - connecting the branches of systems neuroscience. Front Syst Neurosci. 2008 Nov 24;2:4.

L. F. Abbott and P. Dayan, "The Effect of Correlated Variability on the Accuracy of a Population Code," in Neural Computation, vol. 11, no. 1, pp. 91-101, 1 Jan. 1999.

Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci. 2006 Sep;10(9):424-30.

PHILOSOPHY OF THE LIFE SCIENCES

THE METAPHYSICS OF MECHANISMS: AN ONTIC STRUCTURAL REALIST PERSPECTIVE

Yihan Jiang

University of Leeds, Leeds, United Kingdom

Leading metaphysical accounts of mechanisms are based on an ontology of objects or entities (e.g., Glennan 2017, Krickel 2018). These accounts face two challenges: (1) Explain how the metaphysics of mechanisms is related to fundamental metaphysics, which is (presumably) given by fundamental theories of physics. (2) Explain how the entities involved in mechanisms are individuated, and how their boundaries are drawn. The first is pressing in the light of ontological structuralist metaphysics of physics, according to which the world is not fundamentally made up of objects, but rather structures represented by the mathematical equations of our fundamental theories, objects being 'reconceptualised' accordingly (Ladyman and Ross 2007, French 2014). The second problem is pressing since many have pointed out that the way scientists individuate entities and draw their boundaries is related to their purposes and interests, which further arguably leads to an anti-realist position according to which well-delineated entities do not exist in nature. (Bechtel, 2015) I argue these two challenges motivate an ontic structural realist metaphysics of mechanisms.

First, I introduce two leading metaphysical accounts of mechanisms. Both accounts need to answer the so-called bottoming-out question which asks what grounds causality in the fundamental level of reality. I argue that both accounts fail to answer this question because their fundamental ontology is based on objects or entities, which is problematic in the light of philosophy of physics. (Ladyman and Ross 2007) I argue that this motivates building the metaphysics of mechanism bottom-up, that is, to accommodate mechanisms within a physics-informed metaphysics.

Ontic structural Realism has been seen as the ontological theory most consistent with modern physics. After introducing the main themes of OSR, I present my own view that builds upon Krickel (2018)'s metaphysics of mechanism in terms of Entity-involving Occurrents (EIOs). According to Krickel, mechanisms are composed of entities and activities, and both are ontologically primitive. I argue that from the perspective of OSR we can interpret mechanisms as structures and reconceptualize entities as either 'thin' objects whose identities are given contextually by the relevant mechanisms or merely as heuristic devices that allow us to construct, articulate, or represent the relevant mechanisms.

Finally, I argue that my view can better respond to the anti-realist challenge. Many have pointed out that scientists delineate objects or entities according to their explanatory purposes. Depending on the phenomenon they want to explain, scientists would decompose a system, e.g., a human body, into parts in many ways. Anti-realists therefore argue that objects or entities do not exist independent of human 'perspectives', and it is scientists who impose boundaries around entities. (See Kaiser 2018, p.127 and Bechtel 2015) This raises a serious challenge to both Glennan's and Krickle's accounts, as their realism about mechanisms are based on realism about entities. Thus, the advantage of my account is that (structural) realism about mechanisms is compatible with antirealism regarding entities, since the existence of the former does not presuppose the existence of the latter.

French, S. (2014). The Structure of the World: Metaphysics and Representation. Oxford University Press.

French, S., & Ladyman, J. (2011). In defence of ontic structural realism. In A. Bokulich & P. Bokulich (Eds.), Scientific Structuralism (pp. 25-42). Springer Science+Business Media.

Glennan, S. (2017). The New Mechanical Philosophy. Oxford University Press.

Kaiser, M. I. (2017). The Components and Boundaries of Mechanisms. In S. Glennan & P. Illari (Eds.), The Routledge Handbook of Mechanisms and Mechanical Philosophy. Routledge.

Krickel, B. (2018). The Mechanical World: The Metaphysical Commitments of the New Mechanistic Approach. Springer Verlag.

Ladyman, J., & Ross, D. (2007). Every Thing Must Go: Metaphysics Naturalized. Oxford University Press.

Sense of Beauty and Aesthetic Predisposition in Evolutionary Aesthetic Theorising

Pietro Allegretti

The University of Waikato, Hamilton, New Zealand

In this talk I argue that evolutionary aesthetics would benefit from a refined conceptual framework. I will draw attention to a case study central to the aims of evolutionary aesthetics, that is, study of the phylogenesis of secondary sexual characters. The purpose of this talk is to reestablish Darwin's account of sense of beauty (henceforth, 'SoB') as a legitimate concept of evolutionary and philosophical research, and to draw connections between Darwin's account and the work of anthropologist Ellen Dissanayake on aesthetic predispositions. I will show how this synthesis of Darwin and Dissanayake will contribute to evolutionary debates concerning the development of aesthetic characters.

Darwin's account of SoB focuses on different objects than the main concepts of evolutionary aesthetics, like the property of beauty (or state of being beautiful), or judgments of beauty (a cognitive process). Darwin's main purpose was to provide explanations for the (minimally) sense-based perceptual preferences of various species, and his work constitutes a first phylogenetic reconstruction of the components of these preferences (Darwin 1987). I argue that Darwin's SoB is an adequate first guide for identifying which characteristics, both ornamental and behavioural, are the best candidates for evolutionary aesthetic explanation. I provide a framework for characterising the Darwinian SoB's components: first, there is a formal level, expressed in the display of recurrent transpecific geometrical patterns and rhythmical repetition of movements and sounds. Second, an affective, (proto-)emotional or 'felt' level, expressed in the association of sensations of pleasure with specific perceptive patterns.

Although Darwin provided an account fruitful for the first level, he could not achieve an exhaustive account of the second level. Rather, Darwin hoped for new physiological discoveries (Darwin 1890, p. 334). To complete his wish, I argue that Dissanayake's work fits the bill. Using recent discoveries in ethology and neuroaesthetics, Dissanayake studies the development of innate aesthetic predispositions in humans—these give rise to processes of 'artification'—and she shares with Darwin the idea that these predispositions are grounded in formal patterns, emotional reactions and affective states that are common to different species (Dissanayake 2014).

Dissanayake does not explicitly connect SoB with forms of animal bodies or behaviours, but her explanation of aesthetic predispositions ranges over the displays of animals and neurological triggers of these. She shares with Darwin the hypotheses that perceptual discrimination is associated with emotional arousal and these kinds of associations trigger the so-called instinctual emotions, that is, "the hereditary effect of archaic trains of ideas on the structure of the body, permanently linked to pain and pleasure perceptions" (Bartalesi & Portera 2015, p. 105). One implication of this, I argue, is that artification is not a process exclusive to humans; I will indicate several examples of non-human artification triggered by affective responses based on perceptual discrimination. This in turn may influence theories of human aesthetic evolution.

Finally, I shall conclude by indicating a framework for future research. This will involve identifying the phenotypes able to induce pleasure and attraction in individuals and linking them to the broader evolutionary aesthetic discussion.

Bartalesi, L., & Portera, M. (2015). Beyond the nature-culture dichotomy: A proposal for Evolutionary Aesthetics. Aisthesis. Pratiche, Linguaggi e Saperi Dell'estetico, 8(1), Article 1. https://doi.org/10.13128/Aisthesis-16209

Darwin, C. (1987). Charles Darwin's notebooks, 1836-1844: Geology, transmutation of species, metaphysical enquiries. British Museum Natural History.

Darwin, C. (1890). The Expression of the Emotions in Man and Animals (J. Cain & S. Messenger, Eds.; 2nd edition). Penguin Group.

Dissanayake, E. (2014). A Bona Fide Ethological View of Art: The Artification Hypothesis. Hanse Studies, BIS-Verlag Der Carl von Ossietzky Universität Oldenburg, 10, 43–62.

How to Explain the Molecular Make-Up of Life on Earth?

Philipp Spillmann

University of Cambridge, Cambridge, United Kingdom

This paper presents a novel contribution to the philosophy of astrobiology. An unresolved problem in astrobiology is whether the molecular basis of life on Earth is typical for life in the universe. Assumptions about the molecular basis of life on Earth inform astrobiological research in manifold ways (Walker et al. 2018). Yet, scholars often emphasize that it is unclear how reliable these assumptions are when applied to extraterrestrial life (Sterelny 2005, Powell & Mariscal 2015).

According to recent philosophical work, the crucial challenge is to explain why the molecular basis of life on Earth is characterized by an incredibly small subset of physically possible building blocks. According to Simon Conway Morris ('SCM') (2003), Earth-life's molecular make-up is best explained as a series of highly favorable solutions to unavoidable evolutionary problems. Arguably, as only a few possible solutions suffice cumulatively to solve all these problems, we have good reasons to believe that Earth-life's biochemical basis is frequently replicated across the universe. In contrast, Cleland (2019) argues that the molecular make-up of life on Earth is best explained as a frozen accident in prebiotic history. Arguably, we therefore have good reasons to believe that no more than a few of Earth-life's basic molecular constituents are frequently replicated across the universe – if any. I argue that neither of these explanations suffice to establish their desired conclusions. In particular, I defend three claims:

(1) The explanation from favorable solutions and the explanation from frozen accidents do not necessarily constitute 'true rivals': Many of Earth-life's basic molecular constituents could constitute favorable solutions and still be frozen accidents. As both explanations are compatible with each other, it is unclear what it would take to sufficiently establish SCM's/Cleland's desired conclusions.

(2) Neither explanation suffices to account for all of Earth-life's basic molecular constituents. Both explain the choice for those constituents as evolutionary outcomes. Yet, some of these constituents cannot be explained that way. Hence, even if SCM and Cleland were both right, they couldn't explain all relevant aspects of Earth-life's molecular make-up.

(3) Neither SCM nor Cleland consider whether Earth's physical, chemical, and astronomical properties might be atypical for habitable environments in the universe. However, there are good reasons to believe that this might be the case. If Earth is actually 'environmentally' atypical, then this might introduce atypical selection pressures and/or constraints, both of which can feature in explanations for Earth-life's molecular make-up. Such 'environmental' explanations rival both 'favorable solutions' and 'frozen accidents'. Hence, even if SCM's and Cleland's arguments were valid, they couldn't outright establish their conclusions.

I conclude that determining how frequent the molecular basis of Earth-life might be replicated across the universe requires us to account for a much broader set of potential ‘terrestrial’, historical idiosyncrasies than suggested by SCM and Cleland.

- Powell, R. and Mariscal C., 2015. Convergent Evolution as Natural Experiment: The Tape of Life Reconsidered. Interface Focus 5: 20150040. http://dx.doi.org/10.1098/rsfs.2015.0040
- Sterelny, K., 2005. Another View of Life. Stud Hist Philos Sci, 36(3), pp. 585–593.
- Cleland, C.E., 2019. The Quest for a Universal Theory of Life. Searching for Life as We Don't Know It. Cambridge: Cambridge University Press.
- Conway Morris, S., 2003. Life’s Solution: Inevitable Humans in a Lonely Universe. Cambridge: Cambridge University Press.
- Walker, S.I., Bains, W., Cronin, L., DasSarma, S., Danielache, S., Domagal-Goldman, S., Kacar, B., Kiang, N.Y., Lenardic, A., Reinhard, C.T., Moore, W., Schwieterman, E.W., Shkolnik E.L., and Smith, H.B., 2018. Exoplanet Biosignatures: Future Directions. Astrobiology, 18(6), pp. 779-824.

Minimal Agency

Patrick McGivern

University of Wollongong, Wollongong, Australia

Recently there has been a surge of interest in concepts of agency and their applicability in biology and the cognitive sciences (i.e., Fulda 2017, Lyon 2006, Moreno and Mossio 2015, Okasha 2018, Walsh 2015). These accounts shift the focus of discussion of agency toward a wide variety of organisms including insects, plants, bacteria and other single-celled organisms. This can be seen as a recognition that traditional accounts of agency are unduly biased toward the human case: by focusing on organismic agency more generally, we can investigate agency in more diverse and more fundamental cases.

This paper investigates agency in an even broader sense that includes sub-organismal and non-organismal cases. I call this ***minimal agency***. The focus of minimal agency is on the idea that minimal agents are the locus of causal activity (Wilson 2005). I develop this conception using a variety of examples, including motor proteins, viruses and self-propelled oil-drops. The key features of minimal agency, I suggest, involve ***self-directed activity***, and I argue (i) that this basic conception is useful for understanding similarities in phenomena involving a wide range of systems and (b) that this conception can be used as the basis for understanding subclasses of systems and their behaviours, for instance due to different forms of self-directedness. Importantly, this characterisation does not include any specification of the consequences of activity; minimal agents do things, but questions about the significance of their actions are separate from questions about their activity.

I then consider some possible objections to the concept of minimal agency. For example, Moreno and Mossio (2015) reject the idea that examples such as viruses and self-propelled oil-drops can count as agents. I argue that these arguments rely on question-begging criteria, such as the lack of “a metabolism allowing them to display an inherent capacity to modify the conditions at the system-environment interface” (p.96). This indicates that Moreno and Mossio are beginning with a conception of agency centred on organisms and that more minimal instances of agency will be excluded by definition.

Objections such as Moreno and Mossio’s suggest a deeper problem for philosophical discussions of agency. This involves an apparent impasse that arises in cases where discussions of agency begin from different pre-theoretical assumptions about the requirements of agency, and hence lead to different judgments about which things count as agents and which do not. I argue that we can break this impasse by focusing on the theoretical role(s) that agency plays. Here, I draw on O’Malley (2016)’s defence of viral agency. O’Malley’s discussion highlights the significance of broader theoretical frameworks, such as predatory-prey models, that have developed to describe agential behaviors. Using this approach, I use recent work on active materials to describe several different forms of minimal agency. I argue that these cases qualify as instances of agency for reasons analogous to those suggested by O’Malley, despite the fact that they do not involve organisms. I conclude by arguing that recognising and understanding agency in these non-organismal forms is an important step in understanding more complex forms of agency.

References

Fulda, F.C., 2017. Natural agency: The case of bacterial cognition. Journal of the American Philosophical Association, 3(1), pp.69-90.

Lyon, P., 2006. The biogenic approach to cognition. Cognitive Processing, 7(1), pp.11-29.

Moreno, A., Mossio, M., 2015. Biological Autonomy. Springer.

Okasha, S., 2018. Agents and goals in evolution. OUP.

O'Malley, M.A., 2016. The ecological virus. Studies in History and Philosophy of Science Part C, 59, pp.71-79.

Walsh, D.M., 2015. Organisms, agency, and evolution. CUP.

Wilson, R.A., 2005. Genes and the agents of life. CUP.

THE ATTRACTIONS OF CULTURAL SELECTION

Tim Lewens

University of Cambridge, Cambridge, United Kingdom

The 'Paris' and 'California' Schools of cultural evolution place stresses on two different approaches to cultural change (Sterelny 2017). The California School tends to focus on Cultural Selection, roughly understood as 'a Darwinian process comprising the selective retention of favourable culturally transmitted variants' (Mesoudi et al 2006). Meanwhile, the Paris School instead highlights the importance of what is known as Cultural Attraction: as one prominent statement puts it, 'the cognitive mechanisms producing social transmission…do not in general aim at high-fidelity copying as such…. Cultural stability emerges as the cumulative effect of many nonrandom (i.e., biased) transformations' (Scott-Phillips et al 2018). Some commentators (e.g. Sterelny 2017) have suggested a pluralist approach that recognizes the importance of both attraction and selection: in this talk I argue for a novel approach to how this pluralism should be fashioned. Attraction explanations are not the same as selection explanations: attraction explanations focus on the existence of locally stable configurations of cultural items; rather than on local sorting processes corralling cultural variation. This means that cultural attraction and cultural selection should not be understood as different processes that lie at opposite ends of a continuum; instead, cultural attraction names a set of processes through which cultural selection can sometimes (but not always) be realized. This account of the relationship between cultural selection and cultural attraction has numerous payoffs, including the following:

- It gives rise to a disciplined account of what cultural attraction is that does not make the mistake of equating cultural attraction with cultural epidemiology, and which reins in the overly liberal approach of Claidiere and Sperber (see also Buskell 2017)

- It explains and justifies the scepticism of some cultural attraction theorists regarding the value of the Cultural Price Equation (El Mouden et al 2013; Scott-Phillips et al 2018)

- It highlights the ways in which cultural attraction can play an explanatory role in cumulative cultural evolution (Driscoll 2011)

Buskell, A. (2017) 'What are Cultural Attractors?' Biology and Philosophy 32: 377-394.

Claidière, N., T. Scott-Philipps and D. Sperber (2014) 'How Darwinian is Cultural Evolution?' Philosophical Transactions of the Royal Society B 369: 20130368.

Driscoll, C. (2011) 'Fatal Attraction? Why Sperber's Attractors do not Prevent Cumulative Cultural Evolution' British Journal for the Philosophy of Science 62: 301-322.

El Mouden, C., J.-B. André, O. Morin and D. Nettle (2013) 'Cultural Transmission and the Evolution of Human Behaviour: A General Approach Based on the Price Equation' Journal of Evolutionary Biology 27: 231-241.

Mesoudi, A., A. Whiten and K. Laland (2006) 'Towards a Unified Science of Cultural Evolution' Behavioral and Brain Sciences 29: 329–47.

Scott-Phillips, T., S. Blancke and C. Heintz (2018) 'Four Misunderstandings about Cultural Attraction' Evolutionary Anthropology 27: 162-173.

Sperber, D. (1996) Explaining Culture: A Naturalistic Approach. Oxford: Blackwell.

Sterelny, K. (2017) 'Cultural Evolution in California and Paris' Studies in History and Philosophy of Biological and Biomedical Sciences 62: 42-50.

THE EVOLUTION AND ONTOGENY OF NORMATIVE AGENCY: THE ROLE OF JUVENILE SOCIAL PLAY AS BEHAVIOURAL EXAPTATION WITHIN A RESOURCE-RICH DEVELOPMENTAL NICHE

Christopher Joseph An

University of Edinburgh, Edinburgh, United Kingdom

A commonly held view of the evolutionary emergence of normative action asserts that it originates out of requirements for complex social cooperation (Tomasello 2016; Sterelny 2019). In this paper, I propose an alternative account of its origins that is founded not on features of cooperative goal-oriented behaviour, but on shared activities that necessitate robust discursive-communicative features. This is motivated by influential philosophical views asserting the necessary role of a discursive capacity in normative action (Brandom 1994; Darwall 2006). Although complex cooperation may necessitate some form of communication, I deny that it need be discursive. I also deny the common philosophical conflation between discursive communication and linguistic ability; i.e., I claim that discursiveness does not require the possession of language. Instead, I suggest that discursive capacity contains at least three non-linguistic expressive elements: (1) action primarily focussed on means, intention, and performance rather than on ends, goals, and outcomes (Ingram & Moreno-Romero 2021); (2) shared signaling for expressing and attributing intentions and other folk-psychological resources (Andrews 2015); and (3) sustained co-regulated interaction which informs and shapes the intentional meaning expressed in action (Fogel 1993).

An implication of this discourse-oriented approach to the emergence of normative guidance is that more proximate intentional accounts take explanatory priority and are accordingly not sufficiently exhausted by ultimate (selective) explanatory accounts. That is, the functional complexity or apparent design of the relevant behaviours is not so much explained by adaptive-selective evolutionary factors and is instead explained by the role they play in expressing the relevant intentions within the context of the communicative encounter (Scott-Phillips 2015), though notwithstanding any potential secondary effects they have that bear on adaptive-selective value. I suggest that the best way to explain the functional complexity of these discursive behavioural expressions is in terms of psychological and socio-cultural exaptation, over adaptive biological exaptation (Gould 1991; Gabora & Ganesh 2020; Palao 2021; Arvan 2021), as well as its scaffolding through developmental niche construction, over selective niche construction (Stotz 2017; Griesemer 2016; Robles-Zamora 2021).

I motivate this by offering juvenile social play as a concrete instantiation where this proximate behavioural exaptation can be scaffolded through developmental niche construction. What makes play behaviour relevant is that its expression is not primarily driven by its fitness benefits (a "heteronomous" behavioural constraint) but because of its proximate function and motivation of allowing the animal to experience and express its own agency in a more self-motivated and self-directed or "autonomous" manner facilitated by a developmental niche with "surplus resources" (Burghardt 2005). Furthermore, I argue that social play meets the requisite elements for non-linguistic discursiveness through distinctive behavioural markers like self-handicapping, turn-taking, play-signaling, etc. Juvenile social play can thus be seen as a critical formative behaviour system that emerges within resource-rich developmental niches and cultivates a different kind of agent that is not primarily driven by selection pressures. This should suggest that incipient forms of discursiveness that can be found in social play could have a critical formative role in scaffolding robust normative guidance exemplified in autonomous norm-guided agents.

Burghardt, G. M. (2005). The Genesis of Animal Play: Testing the Limits. MIT Press.

Gould, S. J. (1991). Exaptation: A crucial tool for an evolutionary psychology. Journal of social issues, 47(3), 43-65.

Scott-Phillips, T. C. (2015). Meaning in animal and human communication. Animal Cognition, 18(3), 801-805.

Sterelny, K. (2019). Norms and their evolution. In Handbook of Cognitive Archaeology (pp. 375-397). Routledge.

Stotz, K. (2017). Why developmental niche construction is not selective niche construction: and why it matters. Interface focus, 7(5), 20160157.

In search for a generalised view of biological and socio-economic evolution: the Generalised Darwinism programme and Fisher's Fundamental Theorem of Natural Selection

Nicola Bertoldi

Louvain Research Institute for Landscape, Architecture. Built Environment (LAB), Louvain-la-Neuve, Belgium. Centre de Philosophie des Sciences et Sociétés à l'Université Catholique de Louvain (CEFISES), Louvain-la-Neuve, Belgium

Generalised Darwinism (GD) is a currently debated research programme aiming to account for the behaviours of biological and socio-economic forms of organisation by generalising the three "core Darwinian principles" of variation, inheritance and selection (Hodgson & Knudsen 2010). For instance, Thomas Reydon (2021) has recently devised an evaluative framework for assessing GD's explanatory success that expands on Renate Mayntz's formalisation of the concept of "explanatory theory transfer". In this framework, the transfer of theoretical content from one domain to another is explanatorily successful whenever two criteria are fulfilled: first, a "similarity of ***explananda***" criterion, i.e. the possibility of describing relevant ***explananda*** in both domains through "very similar sets of variables, equations, concepts, or narratives"; second, a "similarity of ***explanantia***" criterion, i.e. the possibility of explaining phenomena in both domains by deploying arguments that are "of the same philosophical kind" and "involve very similar ontologies". Based on such criteria, Reydon has concluded that GD does not meet the requirements for successful explanatory theory transfer since its proponents have yet been unable to establish the similarity of biological and socio-economic populations ***qua explananda***, on the one hand, and the similarity of biological and socio-economic principles and models ***qua explanantia***, on the other hand.

Therefore, how would it be possible to reframe GD so as to meet Reydon's objections?

We address this question by comparing GD with a historically significant attempt to generalise Darwin's hypothesis of natural selection (Gayon 1998), i.e. R. A. Fisher's "Fundamental Theorem of Natural Selection" (FTNS), through which Fisher aimed to capture the dynamics of evolutionary change in the context of a comprehensive and abstract selection theory (Fisher 1930). He thus drew inspiration from statistical mechanics and accounting models to establish a formal link between the rate of increase of a biological population's average fitness and its "genetic variance in fitness" (i.e., the portion of fitness variance ascribable to variations in gene frequencies). We argue that Fisher's theorem relies on a specific understanding of populations ***qua*** evolutionary agents, on the one hand, and of the dynamics of selection-driven evolutionary processes, on the other hand, that transcends biology as a particular application domain. This claim thus foregrounds the possibility of comparing GD with the generalised view of evolution arguably embedded in the FTNS. Based on this consideration, our analysis aims to determine whether Fisher's view of evolution is better equipped to capture ontological similarities between biological and socio-economic populations, on the one hand, and dynamical similarities between biological and socio-economic evolutionary processes, on the other hand. To this avail, we highlight formal and material similarities or differences between Fisher's theorem and the two "Fundamental Theorems of Welfare Economics" (Debreu 1983), commonly understood as the mathematical foundations of the general equilibrium theory in economics. By doing so, we assess how the FTNS could define some general conditions for evolutionary equilibrium applicable to populations of biological individuals and groups of socio-economic agents alike, i.e., some general criteria for identifying ontological similarities between populations and dynamical similarities between evolutionary processes in both the biological and the socio-economic realms.

References

Debreu, G. (1983). ***Mathematical Economics: Twenty Papers of Gerard Debreu***. Cambridge: Cambridge University Press.

Fisher, R. A. (1930). ***The Genetical Theory of Natural Selection***. Oxford: Clarendon Press.

Gayon, J. (1998). ***Darwinism's Struggle for Survival. Heredity and the Hypothesis of Natural Selection***. Cambridge: Cambridge University Press.

Hodgson, G. M., and Knudsen, T. (2010). ***Darwin's Conjecture: The Search for General Principles of Social and Economic Evolution***. Chicago: University of Chicago Press.

Reydon, T. A. C. (2021). Generalised Darwinism as modest unification. ***American Philosophical Quarterly***, ***58***(1), 79-94. https://doi.org/10.2307/48600687

Philosophy of the Physical Sciences

Two Forms of Functional Reductionism in Physics

Lorenzo Lorenzetti

University of Bristol, Bristol, United Kingdom

Functionalism is the view that being ***x*** is to play the role of ***x***. In recent years this thesis has grown in importance in the philosophy of physics, where functionalism has been applied to substantially different scientific concepts, such as spacetime, thermodynamic entropy, and classical systems. However, even though a broad range of debates has benefitted from the application of functionalist approaches, the discussion around the correct formulation of functionalism in physics is underdeveloped. The aim of this talk is to assess the traditional functionalist framework and show its limits, and then provide a novel formulation of the functionalist view. Addressing this topic is of crucial importance since a lot of issues hinge on how the view is formulated.

We focus on what is arguably the most important application of this approach, that is functional reduction. To give an example of functional reduction, if we provide a functional characterisation of spacetime in terms of its functional role in general relativity, and we find something in a quantum gravity theory that plays the role of spacetime, we can functionally reduce spacetime to a quantum gravity structure (Lam and Wüthrich 2018, Knox 2019). Functionalism thus allows us to be realist about spacetime even if spacetime turns out to be absent from our fundamental theory of quantum gravity, and it also provides a framework for reduction between the two theories. Considering another example, finding a statistical mechanical realiser that plays the functional role of thermodynamic entropy allows the establishment of a reductive link between thermodynamics and statistical mechanics (Robertson 2020).

Using thermodynamic entropy as a case study, this talk clarifies the meaning of functional reductionism in physics, fleshing out the details of the view beyond its particular applications, in order to define its place with respect to other approaches to reduction, its connection to ontology, and its relation with the syntactic and semantic views of theories. We first critically evaluate the main available framework for functional reduction, defended by Lewis (1970) and Butterfield and Gomes (2020), which we label **Syntactic Functional Reductionism**. This approach is tied to the syntactic view of theories, is committed to a logical characterisation of functional roles, and is embedded within Nagelian reductionism. After highlighting the limits of the view, we propose a new framework, called **Semantic Functional Reductionism**. This novel version of functional reduction adopts a semantic view of theories, spells out functional roles mainly in terms of mathematical roles within the models of theories, and is expressed in terms of the related structuralist approach to reduction. These two approaches to functional reduction also importantly diverge with respect to their ontological implications. While the former is committed to the identification of functionally reduced entities with their realisers, the latter model-based approach is more flexible concerning the ontological aspects of functional reduction.

Providing an alternative to the Lewisian framework is thus vital to the whole functionalist debate, considering the wide-ranging implications yielded by the Lewisian view. Furthermore, from a wider perspective, considering the many successful applications of functionalism, getting a clearer grasp on how it works can help us understand better how theories and their ontologies are related.

1. Butterfield, J. Gomes H. (2020). Functionalism as a species of reduction. http://philsci-archive.pitt.edu/18043/.
2. Knox, E. (2019). Physical relativity from a functionalist perspective. Studies in History and Philosophy of Science
3. Lam, V. Wüthrich, C. (2018). Spacetime is as spacetime does. Studies in History and Philosophy of Science
4. Lewis, D. (1970). How to Define Theoretical Terms. The Journal of Philosophy
5. Robertson, K. (2020) In search of the holy grail: How to reduce the second law of thermodynamics, BJPS

Why go effective?

Michael Miller

University of Toronto, Toronto, Canada

Quantum field theories can be understood as effective theories, that is, as theories with an explicit restriction on their domain of applicability. An important class of quantum field theories must be understood in this way in order to be well-defined. The reasoning underlying this conclusion has historically been bound up with the renormalization problem: some quantum field theories lead to perturbative ultraviolet divergences which necessitate regularization and renormalization. The regularization often takes the form of an ultraviolet cutoff which leads naturally to treating the theory as an effective field theory.

My aim in this talk is to show that the motivation for treating field theories as effective is independent of the renormalization problem understood in this way. That is to say, the presence or absence of perturbative ultraviolet divergence has no bearing on whether or not a quantum field theory must be treated as an effective theory in order to be well-defined. I will argue for this claim by showing that there is a precise sense in which perturbative ultraviolet divergences are mathematical artifacts which result from ill-defined products of distributions. These divergences can be avoided without modifying the physical content of the theory.

It remains true that some field theories must be understood as effective field theories in order to be well-defined. This need stems, however, from non-perturbative ultraviolet problems such as the Landau poles of QED and the Standard Model. The existence of these features is determined by studying the relevant renormalization group flow. In those cases where there is no ultraviolet fixed point and the theory is not asymptotically free or safe, the theory needs to be treated effectively in order for it to be well-defined. Thus, question of whether a theory needs to be treated effectively in order to be well-defined is just the question of whether the theory has a well-defined non-perturbative continuum limit, and this is a question about the nature of the physical interactions posited in the theory which turns out to be independent of the perturbative ultraviolet divergences in the model.

With this clarification in hand, the reasons for construing quantum field theory as effective field theory come into sharper focus. They turn out to have nothing to do with the presence of perturbative ultraviolet divergences, though from a historical perspective it is relatively straightforward to see how the two issues became bound up with one another. I will conclude by arguing that all quantum field theories, not just the non-asymptotically-free and non-asymptotically-safe ones, ought to be treated as effective field theories. This argument amounts to a friendly amendment to recent arguments in favor of interpreting quantum field theory as effective field theory (Fraser 2018, Fraser 2020, Wallace 2006, Wallace 2010, Williams 2019).

Bain, Jonathan (2013). Effective field theories. In Robert Batterman (ed.), The Oxford Handbook of Philosophy of Physics. Oxford University Press.

Fraser, James D. (2020). The Real Problem with Perturbative Quantum Field Theory. British Journal for the Philosophy of Science 71 (2):391-413.

Fraser, James Duncan (2018). Renormalization and the Formulation of Scientific Realism. Philosophy of Science 85 (5):1164-1175.

Huggett, Nick & Weingard, Robert (1995). The renormalisation group and effective field theories. Synthese 102 (1):171 - 194.

Wallace, David (2006). In Defence of Naiveté: The Conceptual Status of Lagrangian Quantum Field Theory. Synthese 151 (1):33-80.

Wallace, David (2010). Taking particle physics seriously: A critique of the algebraic approach to quantum field theory. Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 42 (2):116-125.

Williams, Porter (2019). Scientific Realism Made Effective. British Journal for the Philosophy of Science 70 (1):209-237.

ELIMINATING ELECTRON SELF-REPULSION

Charles Sebens

Caltech, Pasadena, CA, USA

Classical electrodynamics famously faces problems of self-interaction. These problems somehow seem to be resolved in quantum electrodynamics (at least if the theory is formulated properly). This project offers an explanation as to how that feat is accomplished, taking an approach to quantum field theory where fields are viewed as more fundamental than particles (Sebens, 2022).

In classical electrodynamics, point charges face problems of ill-defined self-force and infinite self-energy. One might attempt to alleviate these problems by positing that point charges do not react to their own fields, but then one would miss an important effect: radiation reaction. Charges would emit energy-bearing radiation without losing any energy in the process. Because deleting self-interaction leads to violations of energy conservation, Frisch (2005) has described classical electrodynamics as an inconsistent theory. In a recent talk, Mario Hubert has described the self-interaction problems of classical electrodynamics as parallel to the measurement problem in quantum mechanics. In both cases, foundational problems spur competing proposals as to the laws and ontology of the theory.

The problems of self-interaction for point charges can be avoided by a change to the ontology: replacing point charges with continuous charge distributions. Lazarovici (2018) considers the possibility of addressing the problems of self-interaction in classical electromagnetism via "extended particle models" and criticizes them in part because they require "internal forces of yet unknown origin ('Poincaré stresses') that hold the charge distribution together." This is the problem of self-repulsion. If the electron is modeled classically as a cloud of negative charge, the different parts of that cloud would be expected to repel one another. This self-repulsion would increase the energy of such an electron cloud and would cause the electron to rapidly explode (if there is no force countering the self-repulsion). We do not observe such self-repulsion. As is apparent in quantum chemistry, the ground state energy of an atom or molecule does not include a contribution from electron self-repulsion.

In a classical precursor to quantum electrodynamics (Maxwell-Dirac field theory), the electron can be modeled as a spread-out distribution of charge in the classical Dirac field. This avoids the problems of self-interaction facing point charge models regarding self-force, self-energy, and radiation reaction. However, there remains the problem that the electron would experience self-repulsion. This self-repulsion cannot be eliminated within classical field theory without also losing Coulomb interactions between distinct particles. But, electron self-repulsion can be eliminated from quantum electrodynamics in the Coulomb gauge by fully normal-ordering the Coulomb term in the Hamiltonian. After normal-ordering, the Coulomb term contains pieces describing attraction and repulsion between distinct particles and also pieces describing particle creation and annihilation, but no pieces describing self-repulsion.

Frisch, Mathias. 2005. Inconsistency, Asymmetry, and Non-Locality: A philosophical investigation of classical electrodynamics. Oxford University Press.

Lazarovici, Dustin. 2018. Against Fields. European Journal for Philosophy of Science, 8(2), 145–170.

Sebens, Charles T. 2022. The Fundamentality of Fields. Synthese, 200(5), 380.

Wait, Why Gauge?

Sebastien Rivat

MCMP, LMU Munich, Munich, Germany

Philosophers of physics have spent a considerable amount of effort unpacking the structure of gauge theories. But surprisingly, little attention has been devoted to the question of why we should require our best theories to be locally gauge invariant in the first place, especially in the domain which arguably matters the most, to wit, relativistic quantum field theory (QFT). Agreed: one finds in the literature various justifications appealing to the methodological and theoretical virtues of local gauge invariance (e.g., Belot, 2003, sec. 13). Yet, these justifications usually have little foundational oomph (e.g., mathematical tractability, computational efficiency) or look less compelling once we remember that our best current gauge theories are unlikely to remain empirically reliable at very high energies (e.g., renormalizability). Agreed again: one also finds in the literature some healthy skepticism and more ambitious attempts at explaining the origin of local gauge invariance (e.g., Rovelli, 2014). But even in the latter case, when the plausibly derivative character of local gauge invariance is not just mentioned in passing, we don't really get any non-circular, principled, and convincing explanation as to why we shouldn't include irreducibly locally gauge variant terms in the dynamics of our best theories.

My goal in this talk is to show that there is a specific yet powerful way of looking at relativistic perturbative QFTs in which the principle of local gauge invariance follows from a principle of Lorentz invariance. The argument is mainly based on Steven Weinberg's early treatment of gauge theories in the 1960s (see esp. Weinberg, 1965). I will thus first take the occasion to revisit his work, clarifying a number of important subtleties along the way. Then, I will show how the locally gauge invariant structure of our best theories arises from a demand on their spatio-temporal structure. More precisely, I will make the following claim: for the relevant class of realistic Lagrangian QFT models ***L*** in perturbation theory, if ***L*** is a Lorentz scalar, then ***L*** is locally gauge invariant. For simplicity, I will use pure QED as my main example and only briefly mention how the argument extends to more complicated cases.

The talk will go as follows. I will first explain how to obtain a fundamental constraint on the dynamics of traditional massless fields in perturbation theory by looking at the peculiar space-time transformation properties of massless particles with discrete helicity. Then, I will explain why natural attempts to construct elementary Lorentz covariant fields for such particles are doomed to fail. I will nonetheless argue in a third step that we need such elementary fields to formulate empirically viable relativistic perturbative QFTs, even if these fields are not covariant. I will then show that there is a unique class of such elementary fields allowing us to construct a Lorentz covariant kinetic term out of them. Even more fundamental: I will show that if we require the dynamics constructed out of such fields to be Lorentz invariant, we automatically get local gauge invariance for free. I will conclude with a few interpretative lessons for the philosophy of gauge theories.

Belot, G. (2003). Symmetry and Gauge freedom. ***Studies in History and Philosophy of Science Part B*** 34 (2), 189–225.
Rovelli, C. (2014). Why Gauge? Foundations of Physics 44 (1), 91–104.
Weinberg, S. (1965). Photons and Gravitons in Perturbation Theory: Derivation of Maxwell's and Einstein's Equations. ***Physical Review*** 138 (4B), B988-B1002.

EXPLANATORY DEPTH IN PRIMORDIAL COSMOLOGY:A COMPARATIVE STUDY OF INFLATIONARY AND BOUNCING PARADIGMS

William Wolf[1], Karim Thebault[2]

[1]University of Oxford, Oxford, United Kingdom. [2]University of Bristol, Bristol, United Kingdom

A heated debate has raged in contemporary cosmology regarding the scientific merits of the dominant inflationary paradigm. A small but influential minority of cosmologists have both questioned the justificatory basis for the predominate position of inflation and argued for an alternative paradigm based upon bouncing cosmological models. One axis of this debate relates to the comparison between the two approaches with regards to empirical support both in terms of prediction and accommodation of evidence. Most vividly, dispute along this axis has played out in exchanges between Ijjas et al. (2013, 2014) and Guth et al. (2017, 2014). This aspect has recently received a detailed philosophical analysis from Dawid and McCoy (2021).

Another axis of this debate concerns the relative explanatory merits of the two approaches. In this context, it is worth noting that even the seemingly straight forward explanatory comparison between inflationary explanations and the hot big bang model proves controversial. Inflation was originally motivated by the observation that the hot big bang model involved implausible coincidences which cried out for explanation. However, explication of the basis for the superiority of the inflationary explanation in comparison to the 'fine-tuned' hot big bang models is non-trivial. In particular, as persuasively argued by McCoy (2015), a simple probabilistic framing of the explanatory virtues of inflation in comparison to the fine-tuned hot big bang model suffers from chronic ambiguities in defining probabilistic structure within a modern cosmological context.

A first step towards a more satisfactory, non-probabilistic framing of the explanatory virtues of inflation over the hot big bang is to move away from a reliance on probabilistic notions and, following Maudlin

(2007), focus on the fact that the inflationary explanation is a dynamical one. In this spirit, Azhar and Loeb (2021) have recently argued that dynamical models with initial condition finely-tuning sacrifice explanatory depth and that on this basis the explanatory superiority of inflation over the hot big bang can be established. The notion of depth explored by Azhar and Loeb is a special case of the account from Hitchcock and Woodward (2003), where the explanatory generalization relevant here is initial condition fine-tuning. We will build on this earlier work through a multi-dimensional analysis of the explanatory virtues of inflation in comparison to bouncing cosmologies. In addition to the dimension of explanatory depth in terms of initial condition fine-tuning, we will propose two further dimensions: dynamical fine-tuning and autonomy. We will demonstrate the particular relevance of these dimensions to the inflation vs. bouncing cosmology comparison in the context of dynamical instabilities and the so-called trans-Planckian problem.

Our analysis will not lead us to a verdict with regards to the explanatory superiority of these two rival approaches to cosmology. Rather, we will seek to clarify the relevant terms of the debate, and in doing so, better understand the basis upon which scientists are in fact disagreeing. Furthermore, we will suggest that the different choices with regards to explanatory strategy have direct implications for the heuristics of model building in contemporary cosmology. The nature of the dispute can thus in part be understood in terms of a disagreement over different strategies regarding how best to constrain theoretical practice. Given the heavily unconstrained empirical environment of modern cosmology, such methodological diversity is well justified.

Azhar, Feraz and Loeb, Abraham (2021). “Finely tuned models sacrifice explanatory depth”.

Dawid, Richard and McCoy, CD (2021). “Testability and Viability: Is Inflation-ary Cosmology “Scientific”?”

Hitchcock, Christopher and Woodward, James (2003). “Explanatory generalizations, part II: Plumbing explanatory depth”

McCoy, Casey D (2015). “Does inflation solve the hot big bang model’s fine-tuning problems?”

Maudlin, Tim (2007). The Metaphysics Within Physics

IS SPACETIME CURVED? UNDERDETERMINATION OF RELATIVISTIC GRAVITY THEORIES

Ruward Mulder

Cambridge University, Cambridge, United Kingdom

The existence of teleparallel gravity as an equivalent formulation of general relativity (TEGR) seemingly puts question marks around the interpretation by the scientific realist of curvature as a property of spacetime. Instead of the symmetric Levi-Civita connection, this theory employs an anti-symmetric Weizenböck connection, leading to a non-curved but torsioned spacetime. Indeed, John Earman (1993)---no friend of professional skeptics---takes this as one of just two tough cases of underdetermination in physics (the other one being global properties of empirically indistinguishable spacetimes). Such examples are enough to justify further worries that other theories may be underdetermined by data.

For this reason alone philosophical investigation is warranted, but this has unfortunately has been lacking in the philosophical literature---with notable exceptions in (Lyre & Eynck 2003) and (Knox 2011). The main aim of Knox' paper is to dismantle the underdetermination by stressing the difficulty with `reading off' ontology from formalism, ultimately claiming to have dispelled the threat through a ``relatively liberal attitude to the metaphysical commitments of a theory" (Knox 2011, p. 274). Unfortunately, the way in which we should identify the ontology of a theory under such a liberal metaphysical commitment remains implicit.

I render Knox' arguments more explicit, but contend that they are not conclusive: the liberal attitude entails a metaphysical commitment to a particular understanding of inertial structure, which comes in from outside the theory. Even though Knox rightly draws our attention to the fact that gravitating and otherwise force-free bodies in TEGR follow the autoparallel paths of the Levi-Civita connection and not of the Weitzenböck connection, this fact in itself has no philosophical implications. I contend that there are other philosophical options to evaluate TEGR, corresponding to different standards of inertial motion.

I formulate further worries about TEGR as a serious rival to general relativity---which are not mentioned by Knox but leading to the same conclusion. These are ***Appreciation of Boundaries*** (as articulated by Wolf & Read (forthcoming)), the ***Problem of Operationalisability*** of the torsion tensor and the ***Problem of Visualisability*** of torsion as opposed to curvature. I conclude that Appreciation of Boundaries opens up avenues to break underdetermination locally. Globally, I argue that neither the Problem of Operationalisability nor the Problem of Visualisability provide sufficient relief, as both Problems arise through lack of acquaintance: torsion is just as operationalisable and visualisable as (textbook) curvature, thus (at least up to the boundary term) threatening the realistic interpretation of curvature.

Finally, I consider whether TEGR should be thought of as a geometric theory or a force theory. I come out on the side of the former, arguing that torsion is just as much a geometric property as (textbook) curvature is. Thus, TEGR is just as geometric as general relativity is, even though in TEGR gravitating bodies are pulled off their inertial trajectory. I conclude by investigating options to relax our conceptions of dynamics and geometry as mutually exclusive, taking lessons from Reichenbach's untranslated Appendix to ***Philosophie der Raum-Zeit-Lehre***, as recently interpreted by Giovanelli (2021).

Earman, John (1993). ``Underdetermination, Realism, and Reason." ***Midwest Studies in Philosophy*** **18** (1), pp.~19--38.

Giovanelli, Marco (2021). ``'Geometrization of Physics' vs. 'Physicalization of Geometry.' [...]." In ***Logical Empiricism and the Physical Sciences***. Imprint Routledge. Edited by Sebastian Lutz, Adam Tamas Tuboly. http://philsci-archive.pitt.edu/16831/.

Knox, Eleanor (2011). ``Newton-Cartan theory and teleparallel gravity: The force of a formulation." ***Studies in History and Philosophy of Modern Physics*** **42**, pp. 264--275.

Lyre, Holger & Tim Oliver Eynck (2003). ``Curve it, Gauge it, or Leave it? [...]." ***Journal for General Philosophy of Science*** **34**, pp.~277--303.

Wolf, William & James Read (forthcoming). ``Respecting Boundaries [...]." http://philsci-archive.pitt.edu/21762/.

Back to the Problem of Space

Joshua Eisenthal

Caltech, Pasadena, USA

One of the major problems generated by the discovery of non-Euclidean geometries was the so-called 'problem of space'—the problem of demarcating which mathematical geometries are candidate physical geometries. By the end of the nineteenth century, a broad consensus had developed around a purported solution: only the constant curvature geometries were possible physical geometries. In this paper, I want to consider what difference it might have made if the significance of affine transport (along ***straightest*** lines), independently of metric transport (along ***shortest*** lines), had already been appreciated at this time. This is not entirely fanciful—most if not all of the mathematical and conceptual resources were already available in this period, and it appears to have been a historical accident that a non-metrical notion of affine transport* was only worked out later, specifically following the development of general relativity. Furthermore, as Stachel (2007) has argued, the absence of this notion was largely responsible for the torturous nature of Einstein's path from special to general relativity. With this in mind, I will consider what difference the notion of non-metrical affine transport could have made to the nineteenth century problem of space.

Recall that the nineteenth century consensus was that only the constant curvature geometries were candidate physical geometries. The argument for this conclusion—developed first by Helmholtz and

then made mathematically rigorous by Lie—was based on the idea that the possibility of physical geometry depends on the existence of rigid objects (such as rulers and compasses) which can be moved around without changing their dimensions. If such free mobility was impossible, so the reasoning went, then we would not be able to measure spatial intervals at all. And as only constant curvature geometries have the right congruence structure to capture such free mobility, only those geometries were candidate physical geometries.

This purported solution to the problem of space clearly depends on the notion of metric transport—moving a body such that all of its parts maintain their relative distances. What is out of sight is the significance of affine transport—moving a body such that all of its parts move along parallel lines. Once this is appreciated, however, it quickly becomes evident that the general affine transport of an extended figure is only possible in a flat geometry. If the curvature of space differs from zero, we immediately encounter geodesic deviation. I will argue that, when considering the possibility that space might have a constant curvature, the philosopher-physicists of the nineteenth century could have recognized that the existence of tidal forces would make absolute motion—motion relative to space itself—detectable. Although some might have seen this as indicating that space had to be described by flat Euclidean geometry after all, others might have seen this as pointing to a new way to detect motion relative to the ether.

* More specifically, a non-metrical affine connection; see Stachel (2007) "The Story of Newstein, Or: Is Gravity Just Another Pretty Force?" In ***The Genesis of General Relativity***, Volume 4. Springer.

AN INFERENTIAL-INFORMATION TRANSMISSION ACCOUNT OF OBSERVATION

Sarwar Ahmed

University of Wuppertal, Wuppertal, Germany

The means that scientists employ to observe the world have been changed and/or extended from the naked eye (visible light) to gravitational waves. This indicates two points; first, scientific and technological developments have provided scientists with increasingly more channels to probe the world. Second, the concept of observation, in physical sciences, picked up an epistemic character. In this paper, I argue that the epistemological significance of any observation has two dimensions, namely, the reliability of the observational channel in transferring information and the justifiability of the inferences involved.

In a seminal paper, Dudley Shapere (1982) argued that observation contains epistemological and perceptual dimensions and what has become important in scientific practice is the former. Furthermore, Shapere argues, the epistemological dimension is based on information transmission from the source (object) to the receptor (observer). If the information transmission is without intervention, it is direct observation, otherwise, it becomes inferential (indirect) observation in degrees.

Although this paper is relatively classical, so far there is no systematic criticism of it. Philosophers of science either applied it to historical examples or endorsed a slightly modified version of Shapere's account (Franklin 2017). Most recently, Jamee Elder (2023) criticised Shapere's distinction, and maintained the direct and indirect distinction, however, on another basis.

In this paper, I endorse the idea that information transmission is the epistemological basis for observation, but I deflate the distinction Shapere draws between direct and inferential observations. One can argue against the distinction based on epistemological, historical and practical reasons. However, for the purpose of the talk, I will focus on the epistemological dimension since Shapere is more concerned with epistemological reasons for building this distinction.

I argue that in addition to the fact that every observation is inferential, the inference-free domain (information transmitted from the source to the receptor) is arbitrary. It depends on the point that one assumes to be the end of the observational process. Furthermore, the epistemological significance is based on standards like repeatability, calibration and variation of the information channel and the justifiability of the involved inferences. Therefore, there is no general argument to make Shapere's construal of directness epistemologically superior.

Furthermore, as an alternative, a general account of observation based on information transmission combined with inference to the best explanation is developed that can resolve the problems faced by shapere's account. Since inference to the best explanation is ubiquitous in science (Douven 2022), I argue that it plays an important role in modern observations. To demonstrate that, the observation of the binary black hole systems via gravitational waves (Abbott, B. P., et al. 2016a) is presented as a case study.

Abbott, B. P., et al. (2016a), "Observation of Gravitational Waves from a Binary Black Hole Merger." Physical Review Letters 116 (6): 061102.

Douven, I. (2022), The Art of Abduction, The MIT Press.

Elder, J. (2023), "On the "Direct Detection" of Gravitational Waves" (Under review),

https://www.jameeelder.com/uploads/1/2/1/6/121663585/jan2023_direct_detection_elder.pdf

Franklin, A.D. (2017), "Is Seeing Believing? Observation in Physics." Phys. Perspect. 19, 321–423.

Shapere, D. (1982), "The Concept of Observation in Science and Philosophy." Philosophy of Science, vol. 49, no. 4, pp. 485–525.

COUNTING WHAT COUNTS: SYMMETRY, POSSIBILITY AND INFERENCE

Sean Gryb

University College Groningen, Groningen, Netherlands

The traditional literature on symmetry has focused on the role of symmetry in putting constraints on laws, ontology and empirical content. In this talk, I will argue that considerable new insight can be gained by thinking instead about the role symmetry plays in constraining the ways one counts possibilities. As is well known, many inferences in science involve some sort of evaluation of how likely a particular theoretical hypothesis is given some evidence. Bayesian inferences and contrastive explanations are obvious examples. It follows, therefore, that any principle that has a substantive impact on the kinds of measures that can be put on a particular set of possibilities can have a dramatic effect on the inferential and explanatory structure of a theory. By looking at different examples from physics, I will show that symmetry principles do indeed have exactly this effect.A well known way to use symmetries to constrain the measures of a theory is Jaynes' Maximum Entropy Principle [1]. The purpose of this principle is to arrive at a unique measure for setting priors for some inferential scheme like a Bayesian update. Even in the best case, however, such a principle usually relies on an indifference principle for its justification. On the other hand, many kinds of inferences input a large body of evidence in which case long-term convergence washes out any particular choice of prior. In this case, what matters for the convergence of beliefs is not the prior itself but the topology of the *space* of allowable priors. I will argue that it is the ability of symmetries to change this entire space that allows them to have such a significant impact on the inferential structure of a theory.

To see how this arrises, we will focus on a particular case study brought to light by a recent new proposal [2,3] for explaining the arrow of time. In this proposal, a particular cosmological symmetry called dynamical similarity is used to put restrictions on the natural measures one can use to count possible models of the universe. Concretely, the new symmetry principle rules out the standardly used time-independent measure (called the Liouville measure) and, instead, favours families of time-dependent measures. According to these new time-dependent measures, late-time observers, like those in our present universe, gauge the early state of the universe to be highly typical. This allows for an explanation of the arrow of time that does not require a Past Hypothesis. It further shows how a symmetry principle can dramatically effect certain kinds of explanations — in this case, those about the arrow of time.Further evidence for this mechanism can be found in: anomaly cancellation in the standard model, Lorentz invariance in special relativity, corollaries VI and VII of Newton's Principia, and the singularity theorems of general relativity. We will discuss these briefly subject to time constraints.

1 - Jaynes, Edwin T. "The well-posed problem." Foundations of physics 3, no. 4 (1973): 477-492.

2 - S Gryb, D Sloan. When scale is surplus. Synthese 199 (5-6), 14769-14820.

3 - Barbour, Julian, Tim Koslowski, and Flavio Mercati. "Identification of a gravitational arrow of time." Physical review letters 113, no. 18 (2014): 181101.

The Definition of Spacetime Singularities, Revisited

Samuel Fletcher

University of Minnesota, Twin Cities, Minneapolis, USA

A “singularity” in the context of a physical theory is a location, event, or process where some representationally significant quantity or structure becomes infinite or ill-defined. In general relativity (GR), a ***spacetime*** singularity is a singularity in the metric field, **g**, or one of the significant fields it determines, such as the Riemann curvature. But because in the models of GR (Lorentzian manifolds) **g** is well-defined at every atomic event by definition, there cannot be a location or event where it or one of the structures it determines is ill-defined. Indeed, because **g** represents the structure of durations and lengths, such an event would have to be outside time and space.

The analysis of spacetime singularities concerns the resolution of this puzzle and the detailed account and classification of their types and significance. Aside from its intrinsic interest, it has at least two important implications. Insofar as singularities are taken to betoken some sort of “breakdown” of or “serious physical pathology” in GR’s representational capabilities (Wald 1984; 212, 216), their analysis and consequent prevalence have implications for the bounds of the possibilities that GR represents. Moreover, the “resolution” of these singularities is often taken as a necessary criterion for a satisfactory theory of quantum gravity. In what follows, I outline how mathematical analogies were used to define spacetime singularities, without that definition being completely interpretable in physical terms—at least, until now: I present new results that clarify the stakes involved in accepting the standard definition.

As Geroch (1968) recounts, during the renaissance of general relativity in the 1960s, there were two guiding ideas behind extending the pre-relativistic conception of singularity to **g**. One was that singular spacetimes are precisely those with “missing points,” the points at which **g** would have been defined. Despite this idea’s worthy ambitions, it encountered more problems than successes (Curiel 2021), so the GR community turned to a second idea, that of incomplete curves in spacetime. An incomplete curve is one that must end prematurely—the singular structure obstructs its continuation. This is a proxy for the ill-defined missing points, which are themselves proxies for a singular **g** in the usual sense.

Which curves are incomplete? Earman (1995) has dubbed the notion of ***b-incompleteness***, as formulated by Schmidt (1971), the “semi-official definition.” Schmidt proceeded by analogizing Lorentzian manifolds with Riemannian manifolds. In the latter, all incomplete curves have a bounded parameterization, so Schmidt formulated a “generalized” parameterization that agrees with the usual definition in the Riemannian case but applies also to the Lorentzian. However, as Curiel (2021) complains, adopters of this definition never adequately explained the physical significance of this generalized parameter; the mathematical strategy for defining singularities is thus inconclusive. I make progress on this issue by proving that the b-incomplete curves are precisely those that would have a missing endpoint if they were in Minkowski spacetime. Thus b-incompleteness is more properly understood as a kind of counterfactual missing point criterion for a singular **g**.

Curiel, Erik, 2021, “Singularities and Black Holes”, ***The Stanford Encyclopedia of Philosophy*** (Fall 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2021/entries/spacetime-singularities/>.

Earman, John, 1995, ***Bangs, Crunches, Whimpers and Shrieks: Singularities and Acausalities in Relativistic Spacetimes***, Oxford: Oxford University Press.

Geroch, Robert, 1968, “What Is a Singularity in General Relativity?”, ***Annals of Physics***, 48(3): 526–540.

Schmidt, B.G., 1971, “A New Definition of Singular Points in General Relativity”, ***General Relativity and Gravitation***, 1(3): 269–280

Wald, Robert M., 1984, ***General Relativity***, Chicago: University of Chicago Press.

Everettian Probability as Deterministic Chance

Alexander Franklin

King's College London, London, United Kingdom

In this talk I'll claim that a variant of Hoefer (2021)'s account of probabilities can resolve both the incoherence and the quantitative problems in the Everett interpretation. To do this I'll make use of three principal insights: first (i), that such Humean/Lewisian probabilities supervene in part on a sequence of emergent, branch-relative events, and so are perfectly kosher higher-level chances; second (ii), that the kind of systematisation involved in the Hoefer framework (inherited from Lewis) requires one to take into account structural features of the theory, thus extending Born-rule probabilities into some atypical worlds; third (iii), that the decoherence constraint on emergence entails that the most radically atypical worlds will not, in fact, emerge. The former insight resolves the incoherence problem, as long as some flexibility is granted in the meaning of 'uncertainty', and the latter two insights together ensure that the number of worlds with non-Born rule probabilities is fewer than normally assumed.

(i) It's striking that so few have identified Everettian probabilities as deterministic chances: I'll build on previous work to argue that such probabilities are emergent and, thus fit well into this framework. To do this I'll appeal to theorems claimed in Everett's original work and proved by Barrett (1999), that a typical world (according to the Born measure) will exhibit Born rule frequencies in the long run. Wilhelm (2022) has recently defended a similar approach, however he seeks such probabilities in the fundamental structure of the worlds and he claims that inhabitants of atypical worlds will just have different probabilities, so he is unable to appeal to make use of(ii) and (iii) below. In the talk I'll argue that both these claims of Wilhelm's are in error.

(ii) What to make of atypical worlds? Saunders (2021) suggests that inhabitants of such worlds are simply unlucky. However, the advantage of a Lewisian account of probability is that probabilities do not just supervene just on the sequence of events but are also constrained to fit the structure of the theory. And the Born rule plays such a central feature in the theory, (see e.g. Gleason's theorem) that atypical worlds may still be claimed to feature Born rule probabilities.

(iii) What of the most atypical worlds? In these the frequency of extremely low probability events is such that the considerations of (ii) will not rescue the Born rule. To respond to this I construct an argument inspired by Hanson (2003) to claim that such worlds will not emerge (i.e. will not be larger than the noise of interference) on decoherence accounts of emergence.

Barrett, J. A. (1999). The quantum mechanics of minds and worlds. OUP Oxford.

Hanson, R. (2003). When worlds collide: Quantum probability from observer selection?. Foundations of Physics, 33, 1129-1150.

Hoefer, C. (2019). Chance in the world: A Humean guide to objective chance. Oxford University Press.

Saunders, S. W. (2021). The Everett Interpretation: Probability 1. In The Routledge Companion to Philosophy of Physics (pp. 230-246). Routledge.

Wilhelm, Isaac. (2022). Centering the Everett Interpretation. The Philosophical Quarterly 72.4, 1019-1039.

PHILOSOPHY OF THE SOCIAL SCIENCES

REGULATIVE OPERATIONALISM

Alexander M Mussgnug

University of Edinburgh, Edinburgh, United Kingdom

Within the “Data for Development” movement, researchers increasingly leverage big data and machine learning to measure poverty in the Global South. As of recently, these efforts have progressed from proof-of-concept studies to applications with immediate real-life impact. How do these emerging uses of big data poverty measurements compare to established poverty measurement practice?

On the one hand, developmental economists traditionally employ a pluralist and contextualist stance regarding poverty. Poverty measurements are considered to provide a particular perspective on socioeconomic deprivations, often embedded in a process of political and scholarly deliberation with other sources of information when informing policy efforts. On the other hand, the development and emerging use of machine learning poverty measurements are characterized by a particular operationalist stance. In developing machine learning poverty predictions, researchers often equate the concept of poverty with a specific operationalization of it.

In contrast to some positivist interpretations of operationalism, machine learning researchers are thereby not advocating antirealism about poverty. Rather than a full-fledged semantic thesis, such a position is best characterized as a local methodological commitment, in line with what Uljana Feest (2005) labels methodological operationalism. The operational definitions employed in big data poverty prediction fix the referent of a complex concept and, thereby, often circumvent engagement with the conceptual and normative dimensions of poverty measurement.

However, as machine learning poverty predictions progress from experimental research to their implementation in automated targeting, this methodological commitment not merely determines the referent of a concept for a particular research endeavor but regulates who is considered poor for policy purposes. Immediately applied in policy, operational definitions in the development of machine learning models come to regulate the meaning and application of poverty in the world around us. To stress not merely the methodological nature but, more so, this regulative function of the operationalist stance taken in machine learning poverty prediction and its real-world implications, I introduce the concept of regulative operationalism. I end by highlighting the consequences of this epistemic attitude for the normative dimensions of poverty measurement, emphasizing the close relationship between economic research and policy.

References

Aiken, E., Bellue, S., Karlan, D., Udry, C., & Blumenstock, J. E. (2022). Machine learning and phone data can improve targeting of humanitarian aid. Nature, 603(7903), Article 7903.

https://doi.org/10.1038/s41586-022-04484-9

Feest, U. (2005). Operationism in psychology: What the debate is about, what the debate should be about. Journal of the History of the Behavioral Sciences, 41(2), 131–149. https://doi.org/10.1002/jhbs.20079

Morgan, M. S. (2001). Making Measuring Instruments. History of Political Economy, 33(Suppl_1), 235–251. https://doi.org/10.1215/00182702-33-Suppl_1-235

Vessonen, E. (2021). Conceptual Engineering and Operationalism in Psychology. Synthese, 199(3–4), 10615–10637. https://doi.org/10.1007/s11229-021-03261-x

Reasoning with Models in Thought Experiments: Applying Häggqvist's Template to the Social Sciences

Alexander Linsbichler

Johannes Kepler University Linz, Linz, Austria. University of Vienna, Vienna, Austria. University of Graz, Graz, Austria

A central theme in the methodology of the social sciences is reconstructing how economists and other social scientists reason with models (see Sugden 2002, Morgan 2012, Jhun 2021, and many others). Acknowledging that the role of thought experiments in science as well as the relation between thought experiments and models are contested (see e.g. Reiss 2016, Thoma 2016), this paper provides an account of reasoning with models as centrepieces of thought experiments.

Numerous thought experiments can be reconstructed as one or more arguments (Norton 2004). More specifically, many thought experiments in philosophy and the natural sciences can even be spelled out by one argument template (Häggqvist 2009). This paper substantiates the applicability of Häggqvist's template and of its refinement devised by Linsbichler and da Cunha (2023a, 2023b) to the social sciences. Consequently, at least some instances of thought experimenting in the social sciences are compatible with strict empiricism.

According to Häggqvist and applying his template, many thought experiments can be regimented as consisting of four groups of propositions: i) a counterfactual scenario C which the thought experiment describes as possible, ii) a theory T to be tested, iii) the claim that if scenario C were the case, then a state of affairs W would obtain, and finally iv) the claim that T implies that (if C were the case, then W would not obtain). This account of thought experiments can nicely be integrated with the emerging literature on narratives in economics (see e.g. Morgan & Stapleford 2023), since claim iii) is typically made plausible by narratively tracking the development of the scenario C from its starting position to a state of affairs W.

The paper will illustrate the applicability of the template to reasoning with models in the social sciences by interpreting models as counterfactual scenarios C of thought experiments (see also Suppes 1960). Thus, the model and its animation do not comprise the entire thought experiment and a model by its own is not yet a thought experiment, but models can be employed as centrepieces of thought experiments. Examples include the paradigmatic case of Schelling's checkerboard, reasoning with so-called "imaginary constructions" of Austrian economics, and Otto Neurath's scientific utopianism. Some Neurathian thought experiments prompt conceptual change at the conceptual basis of the discipline of economics in a manner suggested by Kuhn (1964/1977). In particular, Neurath uses thought experiments to advocate irreducibly multidimensional notions of well-being.

Finally, the paper will argue that some instances of reasoning with models in the social sciences do not fit Häggqvist's template without distortion. What is more, some instances of reasoning with models arguably do not even qualify as thought experiments. The paper briefly discusses the status of such likely counterexamples using paradigmatic cases such as the market for lemons and Robinsonades. Yet, Häggqvist's proposal might serve as an impetus to construct additional templates to capture other modes of reasoning with models.

References (Selection)

Häggqvist, S. (2009). A Model for Thought Experiments. Canadian Journal of Philosophy, 39 (1), pp. 55-76.

Linsbichler, A. & da Cunha, I. F. (2023a). Otto Neurath's Scientific Utopianism Revisited. A Refined Model for Utopias in Thought Experiments, Journal for General Philosophy of Science, forthcoming.

Morgan, M. & Stapleford, T. (eds.) (2023). Special Issue on Narratives in the History of Economics, History of Political Economy, forthcoming.

Norton, J. (2004). Why Thought Experiments do not Transcend Empiricism. In C. Hitchcock (ed.), Contemporary Debates in the Philosophy of Science (pp. 44–66), Blackwell.

Reiss, J. (2016). Thought Experiments in Economics and the Role of Coherent Explanations. Studia Metodologiczne, 36, pp. 113-130.

The epistemic and moral ramifications of Epistemic Extractivism for data-sharing in social science research collaborations.

Karl Landström

RSB Lab, Nottingham Trent University, Nottingham, United Kingdom

In this paper I argue that data-sharing constitutes a socio-epistemically and ethically complex practice filled with the potential for significant epistemic benefits, but also for epistemic and ethical wrongs. As academic research has become more data intensive, collaborative data sharing has become more important (Tenopir et. al. 2011). To fully capture the epistemic and ethical complexities of different forms of data sharing I argue that one ought to distinguish between different forms of data-sharing, and primarily between external and internal data-sharing practices. The main focus of the literature on the ethics of data-sharing is data-sharing practices that takes place in interactions between researchers and other actors not affiliated with the research teams they belong to (Alter & Vardigan 2015, Duke & Porter 2015). I call this type of data-sharing external data-sharing. In contrast, internal data-sharing takes place in interactions between different members of the same research-team. I argue that both forms of data-sharing run the risk of becoming epistemically extractivist.

The starting point for both Linda Martín Alcoff's (2022) and Ramon Grosfoguel's (2019) theorisation of epistemic extractivism is that projects that aim to extract resources, be it material or intellectual, have generated knowledge practices, which in turn have generated approaches in epistemology that ought to be reconsidered. Alcoff (2022) argues that such approaches have influenced normative epistemic ideals and presuppositions, particularly in the global North, and that analysis of extractivist epistemologies, and the practices they give rise to, reveal problems with core epistemic ideals. She places the origin of extractivist epistemologies in colonial practices of capitalist extraction and argues that epistemic extractivism comes at both ethical and epistemic costs.

Alcoff (2022) argues that thinking of epistemic extraction as extraction of knowledge is a mistake. Rather, epistemic extraction involves both ontological revisions and new processes of knowledge construction. Assuming that epistemically extractivist practices function as knowledge extraction assumes that individual pieces of knowledge are distinct and separable from the original contexts and relations in which they were created (Alcoff 2022). Rather, epistemic extractivism is a flawed process of meaning-making in which the extractors take themselves to be able to perform the necessary interpretations, analysis and judgements without engaging with the communities and contexts from which the knowledge originated from. In this sense, extractivist epistemologies are highly individualistic. While Alcoff primarily addresses the epistemic problems with epistemic extractivism, other critics, such as Grosfoguel (2019) and Linda Smith (2012), offer arguments against epistemic extraction based on ethical considerations, particularly pertaining to exploitation and undue marginalisation.

To conclude, I draw on existing critiques of social science research collaborations, and a case study of an on-going social science research project to illustrate how data-sharing within social science research collaborations risk becoming epistemically extractivist. Further, I argue for the importance of policies and practices for data production and data-sharing that consider these ethical and epistemic challenges. In doing so, I begin to sketch the outlines of how social science research collaborations can avoid becoming epistemically extractivist.

Alcoff, L. M. (2022). Extractivist epistemologies. Tapuya: Latin American Science, Technology and Society, 5(1).

Alter, G., & Vardigan, M. (2015). Addressing Global Data Sharing Challenges. Journal of Empirical Research on Human Research Ethics. Vol. 10(3) 317–323

Grosfoguel, R. 2019. Epistemic Extractivism. In de Sousa Santos & Meneses (eds.) Knowledges Born in the Struggle. New York: Routledge: 203-218.

Sukarieh, M., & Tannock, S. (2019). Subcontracting academia: alienation, exploitation and disillusionment in the UK overseas Syrian refugee research industry. Antipode, 51(2), 664-680.

Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, et al. (2011). Data Sharing by Scientists: Practices and Perceptions. PLoS ONE 6(6): e21101.

An Eliminativist Account of Validity in Psychology

Oliver Holdsworth

University of Cambridge, Cambridge, United Kingdom

Psychological validity, roughly understood as good measurement of a psychological construct, has a varied definitional history. Taken by some psychologists and philosophers to be a property of the test, and others a property of test inferences, the question of its scope and proper domain is equally controversial and diverse (Feest 2020; Stone 2019; Borsboom and Mellenbergh 2004). Frequent moves to polish up its definition have been met with strong resistance however (Newton and Baird, 2016). Those who argue in favour of a broad notion point to the multiple concerns surrounding measurement (including the value of a test's use in practice). Those who advocate instead for a restricted definition (focusing only on whether a test measures for example), argue that we thereby avoid unwieldiness, in favour of simplicity and practicality.

In the face of this confusion, I argue for an eliminativist account of validity. Multiple validity accounts exist in the literature, but there is neither a desirability nor a possibility to reduce these into a single unitary or dominating concept. I argue that the term 'validity' if left unspecified should refer only to a general amorphous concept, technicality and precision appear when we specify individual validity accounts.

In the paper I will give a taxonomy of these accounts as they exist in the literature. I argue that proposed validity accounts can be profitably split into how each answers the following three questions:

Q1 What subject does the property of validity refer to?

Q2 What relationships are taken as constitutive of validity?

Q3 What is the nature of these relationships?

Answers to ***Q1*** include 'the test', 'an inference' or 'a judgement'. Answers to ***Q2*** include 'the relationship between the target construct and the test scores' or 'the relationship between a test and its uses'. Answers to ***Q3*** include 'the relationship between the construct and scores must be causal' or 'the relationship between the test and its uses must encompass both intended and unintended consequences'. These three questions categorise available validity accounts, and in doing so reveal essential differences and similarities between these accounts, as well as the key and controversial concerns of validity.

As well as presenting this taxonomy, I defend the proposition that there are multiple ways to answer these questions and still produce a viable validity account. This is a necessary defence against those who advocate for a unitary concept of validity, of whatever stripe.

The value of an eliminativist account of validity, and the taxonomy I propose, is two-fold. Firstly, it brings clarity and hopeful resolution to a protracted and unproductive debate over producing a single validity definition. Secondly, it clearly delineates and situates different validity concerns (the precise relationship between a construct and score for example), which in turn facilitates further work on these issues.

Borsboom D, Mellenbergh GJ, van Heerden J. (2004). The concept of validity. ***Psychological Review***. Oct;111(4):1061-71.

Feest, U. (2020). Construct validity in psychological tests – the case of implicit social cognition. ***European Journal of Philosophy of Science*** 10, 4

Newton, P.E & Baird, J (2016) The great validity debate. ***Assessment in Education: Principles, Policy & Practice***, 23:2, 173-177

Stone, C. (2019). A Defense and Definition of Construct Validity in Psychology. ***Philosophy of Science***,86(5), 1250-1261.

Science and Values

When is a Graph Honest? Ethics and Simplification in Science Communication

Corey Dethier

Leibniz University Hannover, Hannover, Germany

In 1999, Geophysical Research Letters published a short paper on temperature changes in the Northern Hemisphere. Included in the paper is a graph that looks vaguely like a hockey stick: a long relatively flat period followed by a sharp uptick. Over the next decade and a half, versions of the "hockey stick graph" became both one of the main tools for communicating the novelty and seriousness of human-induced climate change and one of the main targets of denialist attacks. The furor was exceptional, but the use of the graph itself wasn't. Graphs—and other depictions like diagrams, tables, charts, and pictures—are ubiquitous in the sciences, and particularly important in the context of communicating complex scientific ideas to non-experts. And there are ways of abusing these tools. Fiddle with the scales or axes in the right way, and your graph will appear to show the exact opposite of what the data actually supports. On its face, these practices are dishonest. But applying "honest" and "dishonest" to the use of graphs raises a philosophical problem: it's standard to say that someone is being honest only if they believe that what they're saying is true. But where statements are either true or false, graphs are neither. What then separates honest or otherwise virtuous graphical (pictorial, diagrammatic, etc.) communication from dishonest communication using the same tools?

To answer this question, an analogy is helpful. Traditionally, philosophical discussions of science communication have drawn on analogies to assertion or testimony. My contention is that a better analogy is to map-making: whether and to what degree a map is virtuous depends not only on its content but on how that content is organized, what idealizations and simplifications that map-maker has made in presenting the content. Graphs are similar: elements of the presentation like the choice of scale, the degree of smoothing, which data points are included and which excluded are all essential for the evaluation of graphs as a communicative vehicle. And, I argue, they're also representative in that sense that similar choices matter for science communication even when the vehicle are (written or spoken) sentences.

Adopting this view of science communication requires to reevaluate much of the existing philosophical literature on science communication. For instance, it's common to say that science communication is misleading, dishonest, or untrustworthy insofar as it fails to tell "the whole truth." But excluding information—not telling the whole truth—does not by itself render scientific communication dishonest. On the contrary, excluding information is often an essential part of honest communication, because in the scientific context "the whole truth" is often too complicated to effectively "map out" for audiences with limited amounts of time and background knowledge. The tradeoffs that must be made to account for the audience's limited time and background knowledge force on us a more complicated view in which excluding information can be honest when including that information would only serve to confuse or mislead the audience.

Bibliography:

Dang, Haixin and Liam Kofi Bright (2021). Scientific Conclusions Need Not Be Accurate, Justified, or Believed by their Authors. Synthese 199.3-4: 8187–203.

Fahnestock, Jeanne (1986). Accommodating Science: The Rhetorical Life of Scientific Facts. Written Communication 3.3: 275–96

Gerken, Mikkel (2022). Scientific Testimony: Its roles in Science and Society. Oxford: Oxford University Press.

Grasswick, Heidi E. (2010). Scientific and Lay Communities: Earning Epistemic Trust through Knowledge Sharing. Synthese 177.3: 387–409.

Irzik, Gürol and Faik Kurtulmus (2019). What Is Epistemic Public Trust in Science? British Journal for the Philosophy of Science 70.4: 1145–66.

In Pursuit of Perils: A Social-Epistemological Case Study of Research Method Development in the Biomedical Sciences

Doohyun Sung

KAIST, Daejeon, Korea, Republic of

In light of diverse value issues in science such as the role of scientific evidence in policy decision making, conflicts of interest in medical research, and the replication crisis, philosophers have debated on how epistemic risks in general ought to be addressed and managed during research. Whereas some philosophers have addressed the broader issue of demarcating legitimate roles of values in science, others have advanced reforms concerning specific procedures designed to govern research activities, e.g., peer review. Despite the wealth of philosophical insights and promising proposals, philosophers have just begun to explore one of the most essential contexts of scientific inquiry: research method development. As demonstrated by well-established policy literature (e.g., OECD [2005]), research method development is a critical scientific endeavor which determines the foundation of scientific research against the background of complex relations between epistemic, ethical, and policy values. This is particularly the case for scientific domains whose primary research methodologies remain open to epistemic and ethical controversies (e.g., animal experimentation and in vitro modeling in toxicology).

Based on an ethnographic case study of organoid research at government research institutes, I advance the following theses regarding the management of epistemic risks during research method development. First, as a process of amalgamating novel or existing scientific knowledge and practices into a coherent system (Chang [2014]), method development is distinct from other genres of research (e.g., hypothesis or model-driven research) due to its prioritization of procedural protocol development, proof of concept research (Elliott [2021]), and standardization. By illuminating the epistemic rationales and institutional procedures which reflect such a distinctiveness, I aim to contribute to practice-oriented epistemology of science (Chang [2014]). Second, due to epistemic, ethical, and institutional factors (Elliott [2021]; Laffont and Martimort [2002]), there exist at least two epistemic risks unique to research method development: epistemic siloing (i.e., the accumulation of research outcomes which fail to constitute a coherent system of practice); and epistemic accommodation (i.e., the accommodation of policy interests at the cost of methodological coherence). Moreover, both risks are particularly difficult to address in the face of incentive issues and epistemic asymmetries (e.g., asymmetries of information and expertise) specific to delegatory or principal-agent relations (Laffont and Martimort [2002]). Finally, managing the two risks requires in-situ engagement with the political economy of research and that ethnography is a promising apparatus for this purpose. By drawing from my case study, works in empirical philosophy of science, and the technical literature on organoids, I articulate the advantages and challenges of conducting "cognitive ethnography" (Nersessian and MacLeod [2022]) of strategies for driving and sustaining research method development.

References

Chang, H. [2014]: 'Epistemic Activities and Systems of Practice: Units of Analysis in Philosophy of Science After the Practice Turn', in L. Soler, S. Zwart, M. Lynch and V. Israel-Jost (eds), Science after the Practice turn in the Philosophy, History, and Social Studies of Science, New York: Routledge, pp. 67-79.

Elliott, S. [2021]: 'Proof of Concept Research', Philosophy of Science, **88**, pp. 258-280.

Laffont, J. J. and Martimort, D. [2002]: The Theory of Incentives: The Principal-Agent Model, Princeton, NJ: Princeton University Press.

Nersessian, N. J. and MacLeod, M. [2022]: 'Rethinking Ethnography for Philosophy of Science.', Philosophy of Science, **89**, pp. 721-741.

Organisation for Economic Co-operation and Development (OECD). [2005]: 'Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment', Guidance Document 34, OECD Series on Testing and Assessment, Paris, France: Organisation for Economic Co-operation and Development.

How are researchers told to deal with non-epistemic factors in science? A content analysis of European national documents on research integrity

Jacopo Ambrosj[1], Hugh Desmond[2,3], Kris Dierickx[1]

[1]KU Leuven, Leuven, Belgium. [2]Leibniz Universität Hannover, Hannover, Germany. [3]University of Antwerp, Antwerp, Belgium

Despite the normative character of the debate on values in science little attention has been paid to the norms that researchers are currently asked to follow by documents on research integrity (RI). To fill this gap, we conducted a content analysis of 25 European documents. How do they expect researchers to deal with non-epistemic factors? Do the value-free ideal (VFI) of science play any role in them?

The recommendations on how to deal with non-epistemic factors identified in the documents split in three groups: those expressing a positive attitude, those expressing a negative attitude, and those requiring disclosure. These attitudes are not always mutually exclusive: some factors are sometimes asked to be declared and sometimes to be avoided depending on the context.

In addition to context-sensitivity, we identified two elements of complexity: internal tensions and vagueness. First, different pieces of guidance may be in contrast with each other. In particular, passages expressing VFI-like positions and those stressing the social responsibilities and aims of research. Second, no criteria are provided to evaluate the legitimacy of different factors, when such an evaluation is required. How are researcher supposed to decide whether their interests jeopardize the integrity of research or not?

The tension between endorsing VFI-like positions and stressing the social responsibilities of research may lie in the important functions that the VFI was meant to fulfil. In particular, holding the VFI is supposed to make research objective. To address this, authors of documents on RI may refer to the many different conceptions of objectivity proposed in the literature that not necessarily imply value-freedom.

The VFI is also assigned the function of conveying the image of science as objective to the outside world, promoting trust in research. In this regard, the little evidence currently available suggests that people do not trust researchers more when value-commitments are disclosed. However, in absence of further empirical work, we would suggest not to downplay the capabilities of the public to hold more nuanced views of research than the VFI.

A possible solution to the lack of criteria to evaluate the legitimacy of the influence played by non-epistemic factors would be to standardize and harmonize guidance by adopting a more detailed, step-by-step guidance documents. This detailed oriented guidance may represent a stricter endorsement of the VFI, or be more liberal about the influence of non-epistemic factors. Independently on the content of this detailed oriented guidance, one should consider whether researchers would perceive it as a threat to their autonomy.

By mapping the current guidance on non-epistemic factors and highlighting some of the shortcomings of documents on RI, we hope to encourage policy-makers to thoroughly consider the debate on values in science when drafting new documents on RI. We also suggest that if philosophers of science were to take part into the development of new documents on RI and RI policies in general, philosophy of science could have a greater impact outside its boundaries, as RI policies can put into practice different views on values in science.

Ambrosj, J., Dierickx, K., & Desmond, H. (2023). The Value-Free Ideal of Science: A Useful Fiction? A Review of Non-epistemic Reasons for the Research Integrity Community. Science and Engineering Ethics.

Desmond, H., & Dierickx, K. (2021). Research integrity codes of conduct in Europe: Understanding the divergences. Bioethics.

Holman, B., & Wilholt, T. (2022). The new demarcation problem. Studies in History and Philosophy of Science.

Resnik, D. B., & Elliott, K. C. (2019). Value-Entanglement and the Integrity of Scientific Research. Studies in History and Philosophy of Science

Resnik, D. B., & Elliott, K. C. (2023). Science, Values, and the New Demarcation Problem. Journal for General Philosophy of Science.